

The effect of heterogeneity on decorrelation mechanisms in spiking neural networks: a neuromorphic-hardware study

Thomas Pfeil^{†,1,*}, Jakob Jordan^{2,*}, Tom Tetzlaff², Andreas Grübl¹,
Johannes Schemmel¹, Markus Diesmann^{2,3,4} and Karlheinz Meier¹

¹*Kirchhoff-Institute for Physics, Heidelberg University, Heidelberg, Germany*

²*Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA BRAIN Institute I, Jülich Research Centre, Jülich, Germany*

³*Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*

⁴*Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany*

(Dated: July 6, 2015)

High-level brain function such as memory, classification or reasoning can be realized by means of recurrent networks of simplified model neurons. Analog neuromorphic hardware constitutes a fast and energy efficient substrate for the implementation of such neural computing architectures in technical applications and neuroscientific research. The functional performance of neural networks is often critically dependent on the level of correlations in the neural activity. In finite networks, correlations are typically inevitable due to shared presynaptic input. Recent theoretical studies have shown that inhibitory feedback, abundant in biological neural networks, can actively suppress these shared-input correlations and thereby enable neurons to fire nearly independently. For networks of spiking neurons, the decorrelating effect of inhibitory feedback has so far been explicitly demonstrated only for homogeneous networks of neurons with linear sub-threshold dynamics. Theory, however, suggests that the effect is a general phenomenon, present in any system with sufficient inhibitory feedback, irrespective of the details of the network structure or the neuronal and synaptic properties. Here, we investigate the effect of network heterogeneity on correlations in sparse, random networks of inhibitory neurons with non-linear, conductance-based synapses. Emulations of these networks on the analog neuromorphic hardware system *Spikey* allow us to test the efficiency of decorrelation by inhibitory feedback in the presence of hardware-specific heterogeneities. The configurability of the hardware substrate enables us to modulate the extent of heterogeneity in a systematic manner. We selectively study the effects of shared input and recurrent connections on correlations in membrane potentials and spike trains. Our results confirm that shared-input correlations are actively suppressed by inhibitory feedback also in highly heterogeneous networks exhibiting broad, heavy-tailed firing-rate distributions. In line with former studies, cell heterogeneities reduce shared-input correlations. Overall, however, correlations increase with the level of heterogeneity as a consequence of diminished effective negative feedback.

I. INTRODUCTION

Dynamical systems in nature often exhibit a remarkable degree of diversity, specialization or anticorrelation across their components, despite equalizing factors such as common input or homogeneity in component and interaction parameters. In many cases, these observations can be explained by the effect of negative feedback. Cell differentiation caused by lateral inhibition [1], formation of new species driven by competition [2] or antiferromagnetism [3] constitute just a few examples. In recurrent neuronal networks, inhibitory feedback constitutes a powerful decorrelation mechanism which allows different neurons to respond nearly independently, even if they are driven by largely overlapping local or external inputs [4–6]. Decorrelation by negative feedback hence implements

an efficient form of redundancy reduction. In biological systems, it may serve similar purposes as decorrelation in technical applications, where it is used in data compression (e.g., principal-component analysis [7]), crosstalk reduction (e.g., in digital signal processing [8]), echo suppression (e.g., in acoustics [9]) or random-number generation in hardware [10]. It is tempting to exploit this mechanism in synthetic, neurally inspired architectures such as analog neuromorphic hardware.

Analog neuromorphic hardware mimics properties of biological neural systems using physical models of neurons and synapses (capacitors, for example, emulate insulating cell membranes) [11, 12]. The temporal evolution of the analog circuits represents a solution to the corresponding model equations. In consequence, neural-network emulations on analog neuromorphic hardware are massively parallel, extremely fast and energy efficient. Analog neuromorphic devices are therefore highly attractive as tools for neuroscientific research, e.g., for the investigation of learning on long time scales, and technical applications [13–16]. A biologically inspired neural network (olfactory system of insects) performing rapid on-line data (odor) classification, for example, has recently been successfully implemented on the analog neuromor-

* These authors contributed equally to this study.

† Correspondence: Thomas Pfeil
Kirchhoff-Institute for Physics
Im Neuenheimer Feld 227
69120 Heidelberg, Germany
tel: +49-6221-549813
thomas.pfeil@kip.uni-heidelberg.de

phic hardware system *Spikey* [17, 18]. In this application, decorrelation by inhibition is an essential ingredient to guarantee high classification performance.

For the functional performance of neuronal architectures, the level of correlations between the activities of individual neurons is often pivotal. Whether such correlations are beneficial or not is context dependent. A number of previous studies emphasize a functional benefit of certain types of correlation for encoding/decoding of information in/from populations of neurons [19–21], information transmission [22–24], robustness against noise [25], or gain control of postsynaptic neurons [26]. Other studies argue that positive cross-correlations are detrimental as they decrease the precision or sparseness of population codes [18, 27–30]. Cohen & Maunsell [31], for example, have shown that decreased spike-train correlations in macaque visual area V4 are accompanied by increased behavioral performance in an orientation change-detection task. Depending on the similarity between the trial-averaged responses of different neurons to external stimuli (signal correlation), noise correlations (correlations not explained by signal correlations) can either increase or decrease the amount of information that can be encoded in or decoded from a population of neurons. In populations of neurons with high signal correlation, vanishing or even negative noise correlations are desirable to improve the population code [20].

In finite neural networks, an inevitable source of correlated neural activity is common presynaptic input, shared by multiple postsynaptic neurons. In network models and in-vivo recordings, however, pairwise correlations in the activity of neighboring neurons have been found to be substantially smaller than expected given the amount of shared input [4, 5, 32–35]. In several studies, this observation has been explained by inhibitory coupling. While Ly et al. [36] and Middleton et al. [37] primarily focused on the effect of feedforward inhibition, Renart et al. [4], Wiechert et al. [38], and Tetzlaff et al. [5] attributed the smallness of correlations to an active decorrelation of neural activity by inhibitory feedback. The theory underlying decorrelation by inhibitory feedback suggests the effect to be general: Decorrelation should be observable in any system with sufficiently strong inhibitory feedback, irrespective of the details of the network structure and the cell and synapse properties. For networks of spiking neurons, however, the effect has so far been explicitly demonstrated only for the homogeneous case, where all neurons have identical properties, receive (approximately) the same number of inputs, and, hence, fire at about the same rate [4, 5]. Moreover, the sub-threshold dynamics of individual neurons was assumed to be linear.

Biological neuronal networks typically exhibit broad, heavy-tailed firing-rate distributions [39–45], indicating a high degree of heterogeneity, e.g., in synaptic weights [46–50], in-degrees [51] or time constants [44, 52]. The same holds for neural networks implemented on analog neuromorphic hardware. All analog circuits suffer from device variations caused by unavoidable variability in the man-

ufacturing process. Neurons and synapses implemented in analog neuromorphic hardware therefore exhibit heterogeneous response properties, similar to their biological counterparts [53, 54]. To understand the dynamics and function of recurrent neural networks in both biological and synthetic substrates, it is therefore essential to account for such heterogeneities.

Previous work on recurrent neural networks has shown that heterogeneity in single-neuron properties or connectivity broadens the distribution of firing rates [44, 55] and affects the stability of asynchronous or oscillatory states [51, 56–60]. A number of studies pointed at a potential benefit of heterogeneity for the information-processing capabilities of neural networks [60–71]. The effect of heterogeneity on correlations in the activity of recurrent networks of spiking neurons, however, remains unclear. Padmanabhan & Urban [65] have shown that the responses of a population of unconnected neurons are decorrelated by heterogeneity in the neuronal response properties. These results are supported by the subsequent theoretical analysis in [68]. In the following, we refer to this type of decorrelation by heterogeneity as *feed-forward decorrelation*. It does not account for the effect of the recurrent network dynamics. Active decorrelation due to inhibitory feedback [see above; 4, 5], in contrast, constitutes a very different mechanism. The effect of heterogeneity on this *feedback decorrelation* has lately been studied by Bernacchia & Wang [72] in the framework of a recurrent network of linear firing-rate neurons. In this setup, correlations are suppressed by heterogeneity in the network connectivity (distributions of coupling strengths or random dilution of connectivity). It remains unclear, however, whether this holds true for networks of (nonlinear) spiking neurons.

In this study, we investigate the impact of heterogeneity on input and output correlations in the asynchronous regime of sparse networks of leaky integrate-and-fire (LIF) neurons with conductance-based synapses. Emulation of the networks on the analog neuromorphic hardware system *Spikey* (Figure 1) [17, 73] enable us to investigate the impact of substrate specific properties on the network dynamics. Insights about the interplay between features of the computing substrate and network dynamics are a necessary prerequisite for the development of algorithms that exploit the benefits of analog neuromorphic systems at best.

The configurability of this system [17] enables us to systematically vary the level of heterogeneity, and to disentangle the effects of heterogeneity on feedforward and feedback decorrelation (see above). For simplicity, we focus on purely inhibitory networks, thereby emphasizing that active decorrelation by inhibitory feedback does not rely on a dynamical balance between excitation and inhibition [5, 6]. We show that decorrelation by inhibitory feedback is effective even in highly heterogeneous networks with broad distributions of firing rates (Section III A). Increasing the level of heterogeneity has two effects: Feedforward decorrelation is enhanced, feed-

back decorrelation is impaired. Overall, input and output correlations are increased (Section III B).

Note that results from specific emulations do not directly translate between network emulations on hardware and simulations on conventional computers, because the dynamics, parametrization and interplay of analog circuits is very complex and difficult to reproduce with classical simulations. If simplified models for spatial and temporal variability are considered in software simulations, however, emulation results can be reproduced qualitatively, thereby verifying the design of the hardware system. While our hardware system is designed to physically implement biologically realistic neural algorithms in a fast and energy-efficient way, software simulations are used as a complementary tool to isolate, verify and investigate different hardware features, e.g., spatial and temporal variations. This would be difficult to achieve purely on hardware, due to the limited access and configurability of network parameters. In analogy to the necessity to measure biological tissue to verify assumptions made in Computational Neuroscience, actual emulations on neuromorphic hardware are essential to understand its properties and develop efficient neural algorithms for these devices. The fact that our results hold true for emulations on hardware and simulations with software, and that they can be distilled to simple linear models support the broad relevance of our findings.

II. METHODS

A. Network model

Details on the network, neuron and synapse model are provided in Table I. Parameter values are given in Table II. Briefly: We consider a purely inhibitory, sparse network of N ($N = 192$, unless stated otherwise) LIF neurons with conductance-based synapses. Each neuron receives input from a fixed number $K = 15$ of randomly chosen presynaptic sources, independently of the network size N . Self-connections and multiple connections between neurons are excluded. Resting potentials E_l are set above the firing thresholds Θ (equivalent to applying a constant supra-threshold input current). We thereby ensure autonomous firing in the absence of any further external input. Due to temporal noise, the initial conditions are essentially random.

B. Network emulations on the neuromorphic hardware system *Spikey*

The *Spikey* chip consists of physical models of LIF neurons and conductance-based synapses with exponentially decaying dynamics (Figure 1; for details, see Table I). The emergent dynamics of these physical models represents a solution for the model equations of neurons and synapses in continuous time, in parallel for all units. In

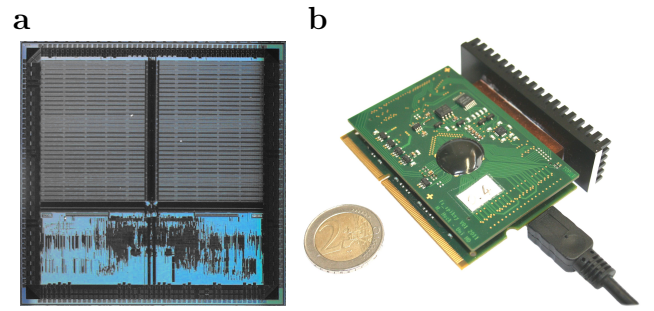


FIG. 1. The neuromorphic hardware system *Spikey*. (a) Photograph of the *Spikey* chip (size $5 \times 5 \text{ mm}^2$). It comprises analog circuits of 384 neurons and 98304 synapses (half of which are accessible for chip version 4 used in this study), is highly configurable and emulates neural-network dynamics by a factor 10^4 faster than biological real-time. (b) Photograph of the *Spikey* system, carrying the *Spikey* chip (covered by a black round seal) and conventional memory. The system is connected to the host computer via USB 2.0, consumes 6 W of power in total and less than 1 nJ per synaptic transmission (see Supplements 1).

contrast, in classical simulations on von-Neumann architectures, model equations are solved by step-wise numerical integration, where parallelization is limited by the available number of virtual processes. To emphasize the difference between simulations using software and simulations using physical models, the term *emulation* is used for the latter [17].

The response properties of physical neurons and synapses vary across the chip due to unavoidable variations in the production process that manifest in a spatially disordered pattern (*fixed-pattern noise*). In contrast to the approximately static fixed-pattern noise, *temporal noise*, including electronic noise and transient experiment conditions (e.g., chip temperature), impairs the reproducibility of emulations. In general, two network emulations with identical configuration and stimulation do not result in identical network activity. Both fixed-pattern and temporal noise need to be taken into account when developing models for analog neuromorphic hardware.

The key features of the *Spikey* chip are the high acceleration and configurability of the analog network implementation. Some network parameters, e.g., synaptic weights and leak conductances, are configurable for each unit, while other parameters are shared for several units (for details see [17]). The hardware system is optimized for spike in- and output and allows to record the membrane potential of one (arbitrarily chosen) neuron with a sampling frequency of 96 MHz in hardware time. Networks on the *Spikey* chip are emulated much faster (approximately 10^4 -fold) than biological real-time, which is a direct consequence of the small capacitances and much higher conductances of VLSI technology compared to biological nervous systems. Due to this high acceleration of the neuromorphic chip, the data bandwidth of the

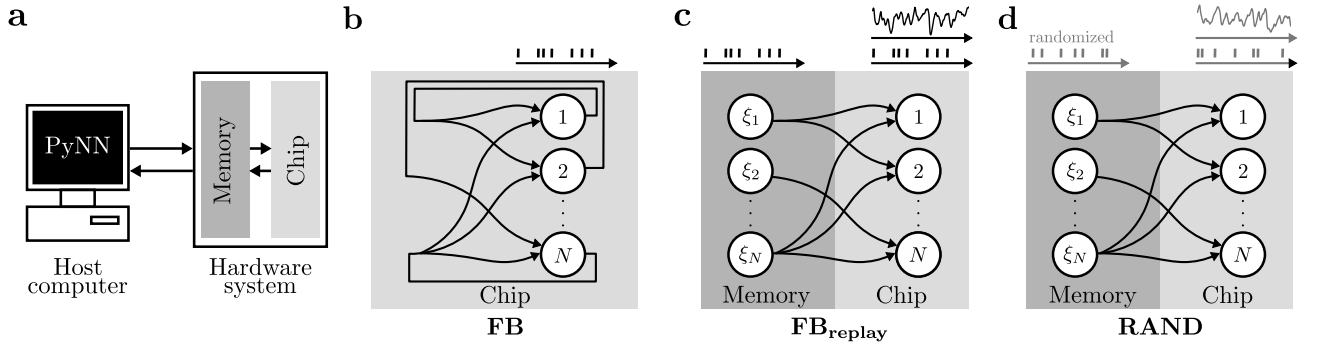


FIG. 2. Experimental setup. **(a)** Data flow of the *Spikey* system. For details see Section II B. **(b)** Network with on-chip feedback connections (FB). Spikes from all neurons are recorded to the local memory. **(c)** Spikes of the FB network in (b) replayed from memory via off-chip spike sources ξ_i to neurons i ($\text{FB}_{\text{replay}}$). Spike times of ξ_i correspond to those recorded from neuron i in (b). Spikes from all neurons or the free membrane potential of one selected neuron are recorded. **(d)** Like (c), but spike times from (b) are randomized for each source (RAND).

connection between the neuromorphic system and the host computer is not sufficient to communicate with the chip in real time. Consequently, input and output spikes (for stimulation and from recordings, respectively) are buffered in a local memory next to the chip. The high acceleration of the *Spikey* chip allows most of the transistors to operate outside of weak inversion, thereby reducing the effect of transistor variations and minimizing fixed-pattern noise.

In contrast to such accelerated systems, most other configurable, analog neuromorphic substrates are designed for real-time emulations at very low power consumption [74–80] and implement fewer, but more complex, neurons [81, 82].

Access to the *Spikey* system is encapsulated by the simulator-independent language PyNN [83, 84], providing a stable and user-friendly interface. PyNN integrates the hardware into the computational neuroscience tool chain and has facilitated the implementation of several network models on the *Spikey* chip [17, 18, 85–87].

On the *Spikey* system, a spiking neural network is emulated as follows (Figure 2a): First, the network described in PyNN is mapped to the *Spikey* chip, i.e., neurons and synapses are allocated and parametrized. Second, input spikes, if available, are prepared on the host computer and transferred to the local memory on the hardware system. Third, the emulation is triggered and available input spikes are generated. Output spikes and membrane data are recorded to local memory. Last, spike and membrane data are transferred to the host computer and scaled back into the biological domain of the PyNN model description.

For consistency with the model description and simplified comparison to the existing literature, all hardware times and all hardware voltages are expressed in terms of the quantities they represent in the neurobiological model, throughout this study.

C. Experimental setup

To differentiate and compare the effects of shared inputs and feedback connections on correlations, we investigate two different emulation scenarios: First, we emulate networks with intact feedback (FB, Figure 2b), and second, the contribution of shared input is isolated by *randomizing* the temporal order of this feedback (RAND, Figure 2d).

In the RAND scenario, the inputs of neurons are decoupled from their outputs. Spatio-temporal correlations in presynaptic spike trains are removed by randomizing the presynaptic spike times.

Input correlations between neurons are measured via their *free membrane potential*, i.e., the membrane potential with disabled spiking mechanism (technically, the threshold is set very high). Because membrane potential traces can be recorded in the hardware only one at a time, traces are obtained consecutively, while repeatedly *replaying* the previously recorded activity of the FB network to a population of unconnected neurons of equal size. We keep the connectivity the same, and hence each neuron receives the same number of spikes as in the recurrent network during the whole emulation, either without ($\text{FB}_{\text{replay}}$, Figure 2c) or with randomization of presynaptic spike times (RAND), respectively. To preserve the fixed pattern of variability of synaptic weights in hardware, the same hardware synapses are used for each connection in both scenarios. If network dynamics were reproduced perfectly, membrane potential traces and spike times would be identical in the FB and $\text{FB}_{\text{replay}}$ cases (see also Section II D).

Drawing two different network realizations (i.e., the connectivity matrix) results in the allocation of different hardware synapses, and, due to fixed-pattern noise, in different values of synaptic weights. To average over this variability, throughout this study, emulation results are averaged over $M = 100$ network realizations, if not stated otherwise.

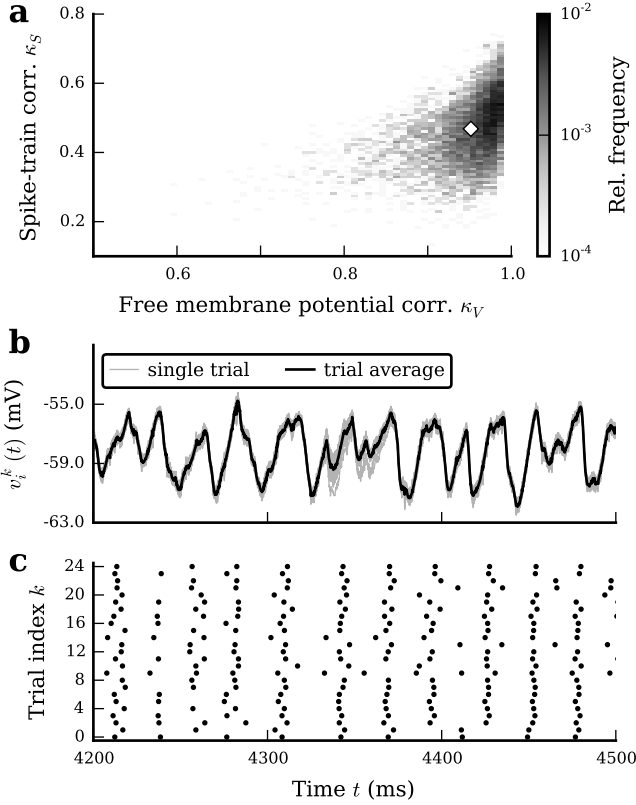


FIG. 3. Reproducibility of free membrane potentials and spiking activity in the $\text{FB}_{\text{replay}}$ case. (a) Low-frequency coherence κ_V and κ_S of free membrane potentials $v_i^k(t)$ and $v_i^l(t)$ and binned spike trains $s_i^k(t)$ and $s_i^l(t)$, respectively, for each neuron i averaged over $L = 25$ trials k, l with $k \neq l$, for $M = 50$ different network realizations. The diamond marks the average across all neurons i and M network realizations. (b) Free single-trial membrane potentials $v_i^k(t)$ (gray) and average over trials $\frac{1}{L} \sum_{k=1}^L v_i^k(t)$ (black) and (c) spike density $\xi_i(t)$ of a single neuron i for $L = 25$ identical trials. The selected neuron i has membrane potential coherence and spike train coherence closest to the diamond in (a).

D. Reproducibility of hardware emulations

Due to the fact that the initial conditions of the recurrent network on hardware are undefined, consecutive emulations of the FB network result in different network activities. In the RAND and $\text{FB}_{\text{replay}}$ case, however, the input of neurons is decoupled from their output. Although unavoidable temporal noise is present, the system's trajectory returns to the trajectory of the previously recorded FB case. A certain degree of reproducibility is required for two reasons: First, the investigated effect of decorrelation by inhibitory feedback requires a precise relation between spike input and output. Thus our method of replacing the feedback loop by replay is only valid if temporal noise does not substantially corrupt this relationship. Second, to record the membrane potentials of all neurons, as if recorded at once, neuron

dynamics have to be reasonably similar in consecutive emulations.

We measure the reproducibility of neuron dynamics by comparing consecutive emulations with identical configuration, i.e., connectivity and stimulation. For this purpose the spiking activity of a FB network is first recorded (Figure 2b) and then repeatedly replayed (Figure 2c). Reproducibility is quantified by the correlations (κ_X in Table III) of free membrane potential traces and output spike trains obtained for individual neurons in $L = 25$ different trials.

Free membrane potentials are reproduced quite well, while spike trains show larger deviations across trials (Figure 3). Small deviations in the membrane potential (Figure 3b) are amplified by the thresholding procedure [38, 88, 89] and can lead to large differences between spike trains (Figure 3c). Consequently, measures based on data of several consecutive replays are more precise for membrane potentials than for spike trains. Nevertheless, results have to be interpreted with care in both cases.

E. Calibration

The heterogeneity of the *Spikey* hardware is adjusted by calibrating the leak conductance¹ for each individual neuron, compensating for fixed-pattern noise of neuron parameters. To this end, a population of unconnected neurons is driven by a supra-threshold constant current influx and the time-averaged population activity \bar{r} is measured. Then, we applied the bisection method [90] to adjust the leak conductance g_l of each neuron, such that the neuron's firing rate matches the target rate \bar{r} . This results in calibration values b for the leak conductance $g_l = g_{l,0}(1 + b)$, where $g_{l,0}$ is the leak conductance before calibration. Because emulations on hardware are not perfectly reproducible, more precise calibration was achieved by evaluating the median over 25 identically configured trials instead of single trials. Furthermore, the bisection method was modified for noisy systems (for details, see Supplements 2).

Intermediate calibration states are obtained by linearly scaling the full calibration:

$$g_l = g_{l,0}(1 + (1 - a)b) \quad . \quad (1)$$

The heterogeneity a is chosen in $[0, 1]$ for calibrations between the uncalibrated ($a = 1$) and calibrated state ($a = 0$). In the following, the fully calibrated chip ($a = 0$) is used, if not stated otherwise.

This calibration substantially narrows the distribution of firing rates compared to the uncalibrated state (Figure 4). With respect to the stationary firing rate, variability on the neuron level is reduced from 38% to 2%.

¹ since capacitances and potentials can not be configured individually for each hardware neuron [17]

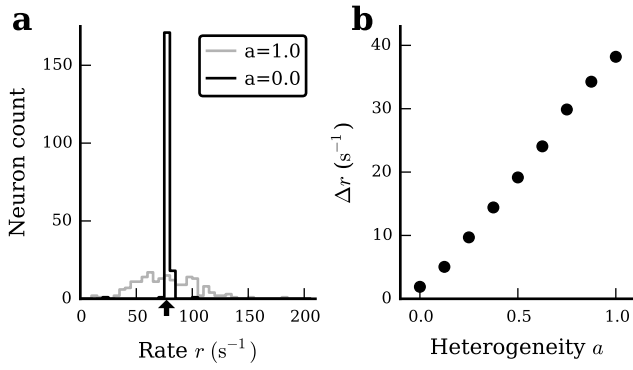


FIG. 4. Calibration of the *Spikey* chip. **(a)** Histogram of firing rates r for a population of unconnected neurons with supra-threshold input currents, before (gray) and after (black) calibration, each neuron averaged over $L = 100$ trials. The arrow denotes the target rate \bar{r} . **(b)** Difference $\Delta r = r_{P75} - r_{P25}$ of 75th and 25th percentile of the histograms in (a), as a function of network heterogeneity a (Equation 1). The mean firing rate over all values of a is $(78.1 \pm 0.7) \text{ s}^{-1}$.

Note that after this calibration procedure the hardware network is still not homogeneous. In addition to remaining variations in neuron parameters, synaptic parameters have a significant variation [18, 91].

F. Correlation measures

In the following, we introduce definitions used to analyze the recorded data. For clarity, all relevant equations and their parametrization are listed in Table III and IV, respectively.

We quantify correlations of membrane potentials $v_i(t)$ and spike trains $s_i(t)$ by the population-averaged *low-frequency coherence* κ_V and κ_S , respectively. At frequency zero, the coherence corresponds to the normalized integral of the cross-covariance function, i.e., it measures correlations on all time scales. We define the low-frequency coherence κ_X , with $X \in \{S, V\}$, to be the average coherence over a frequency interval from 0.1 to 20 Hz. In this interval, the suppression of population-rate fluctuations in recurrent networks due to inhibitory feedback is most pronounced, and the coherence is approximately constant. Before calculating the coherence, we convolve the power- and cross-spectra with a rectangular window to average out random fluctuations. This measure, or a variant of it, is commonly used in the neuroscientific literature [4, 5, 68, 89, 92–94]. We use the terms low-frequency coherence and correlation interchangeably.

Throughout this study, the term *input correlations* is used for correlations between free membrane potentials, and *output correlations* for correlations between spike trains. *Shared-input correlations* are membrane potential correlations that are exclusively caused by overlapping presynaptic sources, ignoring possible correlations in the presynaptic activity. The average pairwise shared-

input correlations in a homogeneous network are of the size of the connectivity [5]:

$$\kappa_V = K/N. \quad (2)$$

We assess the significance of correlations by comparing the results from emulations to correlations in surrogate data, in which we removed spatial correlations. For every neuron, we randomly shuffled bins of the membrane potential trace, and assigned a new timestamp uniformly drawn from the emulation interval to every spike. We thereby remove all spatio-temporal correlations between neurons recorded in parallel. By this procedure we create 100 surrogate trials, across which we calculate the average correlations and the standard error.

To quantify fluctuations in the population activity \bar{s} (Figure 5a–c, horizontal histograms) we compute the power spectrum $\bar{A}(f)$ of the population activity (Figure 5e), which we scale with the duration T of the emulation. Consequently, the population power spectrum $\bar{A}(f)$, scaled by the population size, coincides with the time-averaged population activity \bar{r} for high frequencies: $\lim_{f \rightarrow \infty} \frac{1}{N} \bar{A}(f) = \bar{r}$ [34].

As a measure of pairwise correlations in the time domain (Figure 5d), we compute the population-averaged cross-correlation function $c(\tau)$ by Fourier transforming the population-averaged cross-spectrum $C(f)$ to time domain.

III. RESULTS

In this study, we investigate the roles of shared input, feedback and heterogeneity on input and output correlations in random, sparse networks of inhibitory LIF neurons with conductance-based synapses (Table I), implemented on the analog neuromorphic hardware chip *Spikey* (Figure 1). Similarly to [5], we separate the contributions of shared input and feedback by studying different network scenarios (Figure 2): In the FB case, we emulate the recurrent network with intact feedback loop (Figure 2b) and record its spiking activity (Figure 5a). In the FB_{replay} case (Figure 2c), the feedback loop is cut and replaced by the activity recorded in the FB network. Ideally, the input to each neuron in the FB_{replay} case should be identical to the input of the corresponding neuron in the FB network. As the replay of spikes and the resulting postsynaptic currents and membrane potentials are not perfectly reproducible on the *Spikey* chip, the neural responses in the FB and in the FB_{replay} scenario are slightly different (compare Figure 5 a to b). In the RAND case (Figures 2d and 5c), we use the same setup as in the FB_{replay} case. However, the spike times in each presynaptic spike train are randomized. While the average presynaptic firing rates and the shared-input structure are exactly preserved in this scenario, the spatio-temporal correlations in the presynaptic spiking activity are destroyed.

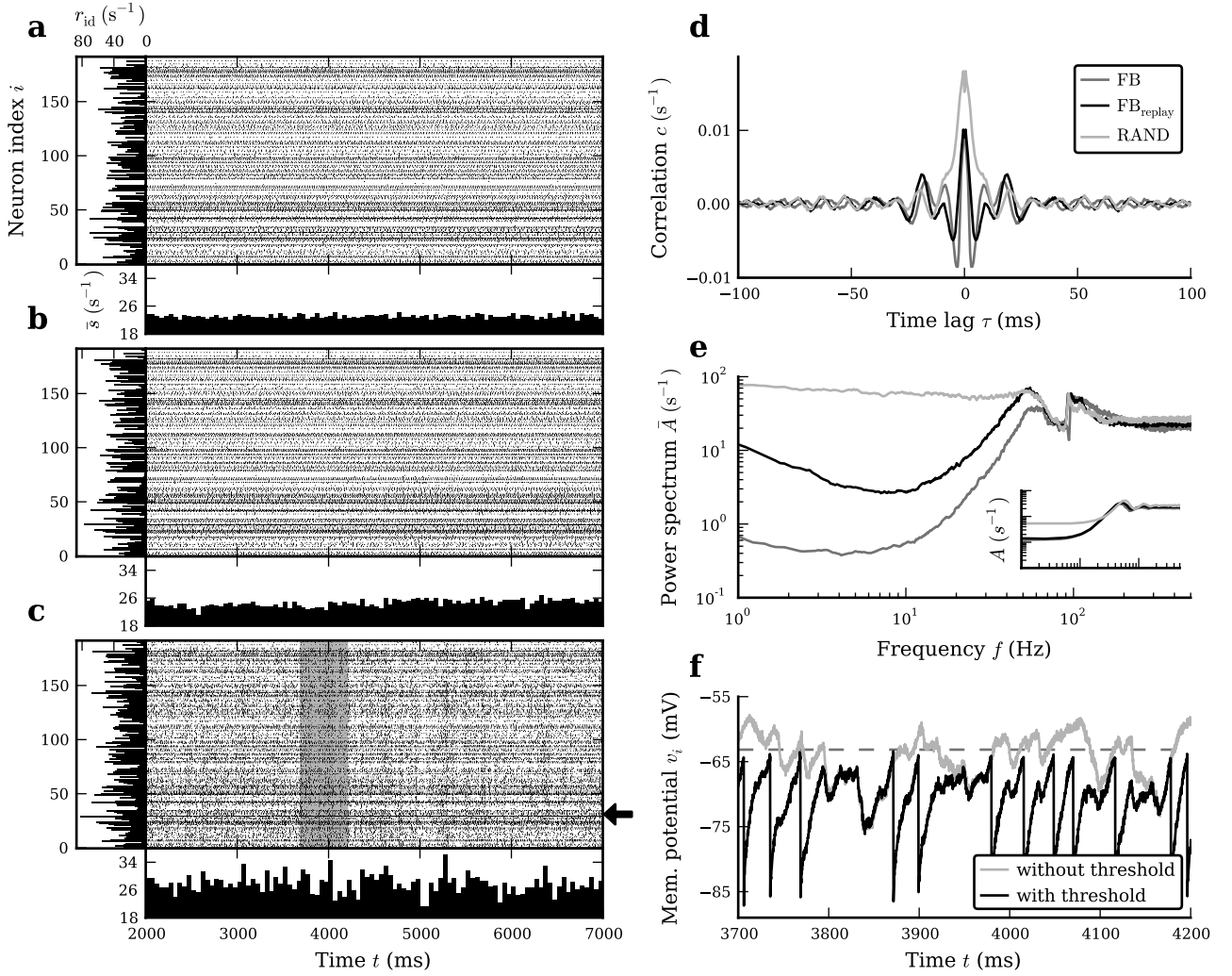


FIG. 5. Spiking and membrane-potential activity in a random inhibitory network of LIF neurons with intact and cut feedback loop. (a–c) Spiking activity (raster plots), population activity $\bar{s}(t)$ (horizontal histograms; bin size 50 ms) and time-averaged single-neuron firing rates r_{id} (vertical histograms) in the network with intact feedback (a) and for cases where the feedback loop is cut (b and c). (a) Intact recurrent network (FB scenario). (b) Population of mutually unconnected neurons receiving identical input spike trains as in (a) (FB_{replay} scenario). (c) As in (b), but after randomization of presynaptic spike times (RAND scenario). (d and e) Population-averaged cross-correlation functions $c(\tau)$ (after offset subtraction) of pairs of spike trains (d) and power spectra $\bar{A}(f)$ (e; log-log representation) of the population activity $\bar{s}(t)$ (cf. horizontal histograms in (a–c)) for the FB (dark gray), FB_{replay} (black) and RAND scenario (light gray). Inset in (e): Population-averaged power spectra $A(f)$ of individual single-cell spike trains (same scales as in main panel). Correlation functions and spectra are averaged across $M = 100$ network realizations. (f) Membrane potential of a neuron in the RAND scenario (with firing rate of 29.0 s^{-1} close to population average of 27.8 s^{-1} ; see black arrow in (c)) with intact (black curve) and removed threshold (gray curve; *free membrane potential*). The threshold potential is marked by the horizontal dashed line. The time frame corresponds to the gray-shaded region in (c).

Using this setup, we first demonstrate in Section III A that active decorrelation by inhibitory feedback [4, 5] is effective in heterogeneous networks with conductance-based synapses over a range of different network sizes. In Section III B, we show that decreasing the level of heterogeneity by calibration of hardware neurons leads to an enhancement of this active decorrelation and thereby to a decrease in input and output correlations.

A. Decorrelation by inhibitory feedback

The time-averaged population activities in the FB, FB_{replay} and RAND scenarios are roughly identical (Figure 5a–c; see also high-frequency power in Figure 5e). In the FB and FB_{replay} scenario, fluctuations in the population-averaged activity are small (horizontal histograms in Figure 5a and b). The removal of spatial and temporal correlations in the presynaptic spike trains

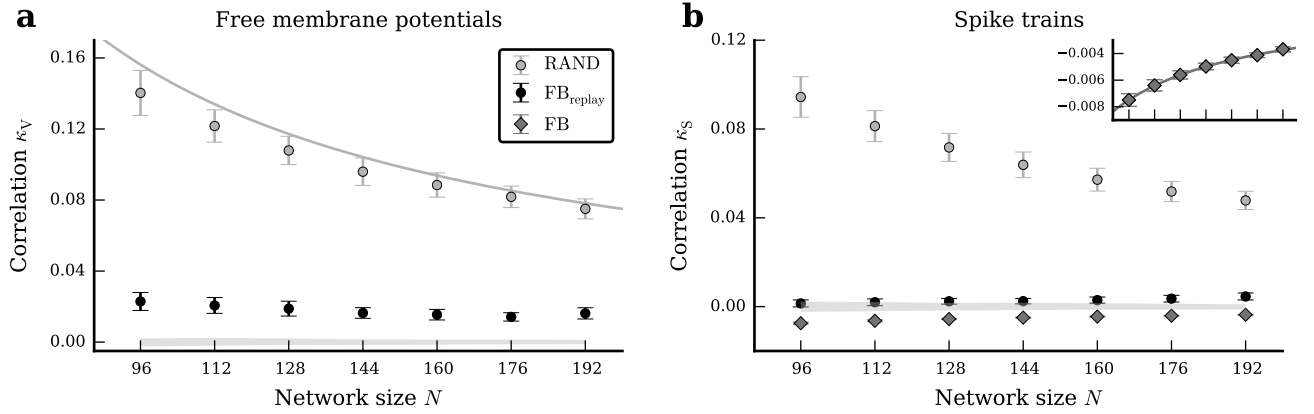


FIG. 6. Dependence of population-averaged input correlations (a) and spike-train correlations (b) on the network size N , for the intact network (FB, dark gray diamonds), the FB_{replay} (black circles) and the RAND (light gray circles) case (fixed in-degree $K = 15$). Symbols and error bars denote mean and one standard deviation, respectively, across $M = 100$ network realizations (error bars are partly covered by markers). Gray curve in (a) depicts shared-input correlations in a homogeneous network (Equation 2). The inset in (b) shows a magnified view of the spike-train correlations in the FB case (dark gray diamonds) with a power-law fit $\sim N^{-1}$ (dark gray curve). The light gray horizontal band represents mean \pm three standard deviations of correlations in surrogate data, in which spatial correlations were removed. Note that free membrane potentials cannot be recorded in the FB case (see Section II). Hence, there are no gray diamonds in (a).

in the RAND case leads to a significant increase in the fluctuations of the population-averaged response activity (horizontal histograms in Figure 5c). At low frequencies (≤ 20 Hz), the population-rate power in the FB and in the RAND case differs by almost two orders of magnitudes (dark and light gray curves in Figure 5e). This increase in low-frequency fluctuations in the RAND case is mainly caused by an increase in pairwise correlations in the spiking activity (Figure 5d; the power spectra of individual spike trains [inset in Figure 5e] are only marginally affected by a randomization of presynaptic spike times) [5]. In other words, shared-input correlations, i.e., those leading to large spike-train correlations in the RAND scenario, are efficiently suppressed by the feedback loop in the FB case.

On the neuromorphic hardware, the replay of network activity is not perfectly reproducible (Section IID). While the across-trial variability in membrane potentials is small, postsynaptic spikes are dithered on a timescale of approximately 5 ms (Figure 3). In the FB_{replay} case, the suppression of shared-input correlations by correlations in presynaptic spike trains is slightly less efficient as compared to the intact network (FB). The differences in the population-rate power spectra and in the spike-train correlations between the FB_{replay} and RAND case, respectively, are nevertheless substantial (solid black and light gray curves in Figure 5d and e; note the logarithmic scale; for a detailed investigation of spike dither see Supplements 4 and Supplements Figure 7).

In the RAND case, presynaptic spike-train correlations were removed, and hence input (i.e., free-membrane-potential) correlations are exclusively determined by the number of shared presynaptic sources (Equation 2). If

the in-degree K is fixed, input correlations will decrease with network size N (Equation 2, light gray curve and symbols in Figure 6a). In contrast, for purely inhibitory recurrent networks (FB scenario), correlations in presynaptic spike trains are on average significantly smaller than zero (dark gray diamonds in Figure 6b, [5]), and largely cancel the positive contribution from shared-input correlations. Average input correlations are therefore significantly reduced (black symbols in Figure 6a). As both shared-input and spike-train correlations scale with the inverse of the network size (N^{-1} ; light gray curve in Figure 6a and inset in Figure 6b, respectively) [89], this suppression of correlations in the FB (and FB_{replay}) case is observed for all investigated network sizes N . Note that output correlations are negative even though input correlations are positive. This effect is predicted by theory and also observed in linear network models as well as LIF-network simulations on conventional computers (see Figure 9, Supplements 4, Supplements 5 and Section IV).

B. Effect of heterogeneity on decorrelation

In neural networks implemented in analog neuromorphic hardware, neuron and synapse parameters vary significantly across the population of cells (fixed-pattern noise; see Section IIB). For a population of mutually unconnected neurons with distributed parameters, injection of a constant (supra-threshold) input current leads to a distribution of response firing rates (Figure 4). In this study, we consider the width of this firing-rate distribution as a representation of neuron heterogeneity. It is systematically varied by calibration of leak conductances.

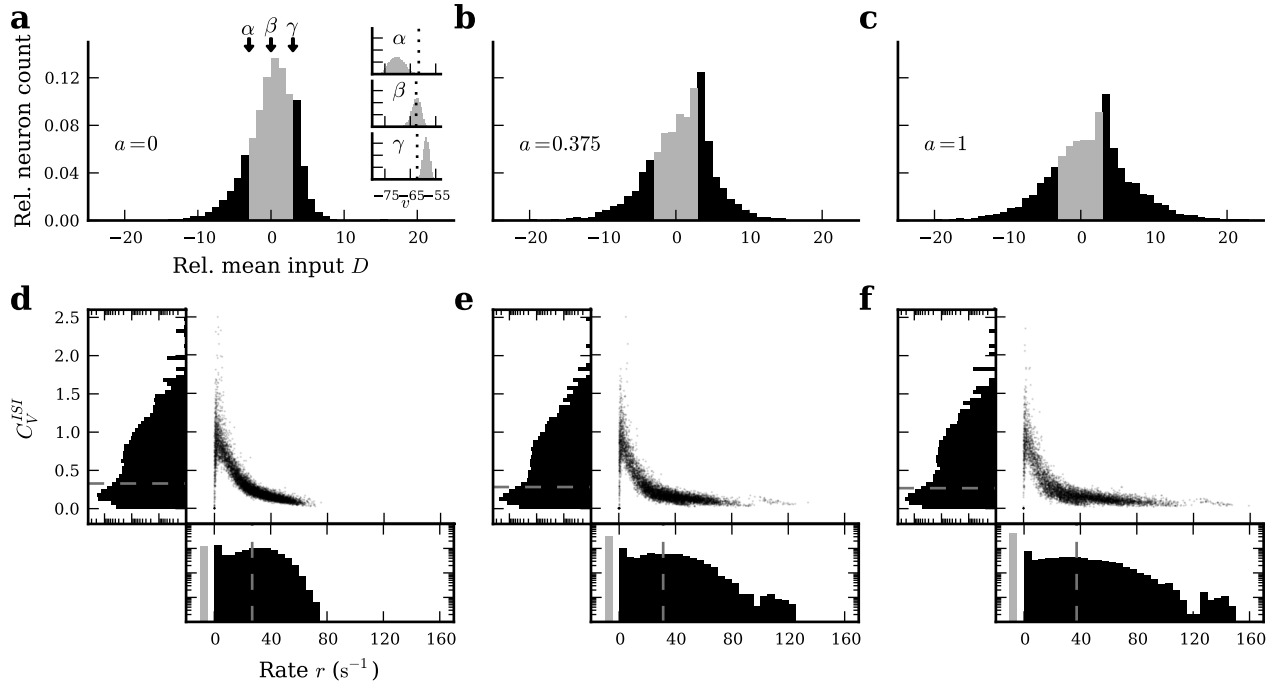


FIG. 7. Modulation of network heterogeneity by leak-conductance calibration (see Section II E). Input (top row) and firing statistics (bottom row) in the intact recurrent networks (FB scenarios) for fully calibrated (a and d; $a = 0$), partially calibrated (b and e; $a = 0.375$) and uncalibrated neurons (c and f; $a = 1$). (a–c) Effect of calibration on input statistics. Distributions of relative mean input $D = (\bar{v} - \Theta)/\sigma(v)$ (distance of time averaged free membrane potential \bar{v} from firing threshold Θ in units of the standard deviation $\sigma(v)$) across the population of neurons. Gray areas in (a), (b) and (c) highlight $[-3, 3]$ intervals, containing 88%, 69% and 60% of the total mass of the distribution, respectively. Inset in (a): Distributions of free membrane potentials v for three neurons α , β and γ with $D = -3$, $D = 0$ and $D = 3$ (arrows in (a)), respectively. Dotted lines mark threshold potentials that may vary due to fixed-pattern noise. (d–f) Effect of calibration on spike-train statistics. Joint (scatter plots) and marginal distributions of single-neuron firing rates r (horizontal histograms; log-linear scale) and coefficients of variation C_V^{ISI} of inter-spike intervals (vertical histograms; log-linear scale). Dashed lines mark mean of firing rate (26.9 s^{-1} , 31.2 s^{-1} , 37.3 s^{-1}) and C_V^{ISI} distributions (0.33, 0.28, 0.27), respectively. Gray bars (bottom panels) represent fractions of silent neurons. Data obtained from $M = 50$ different network realizations.

The extent of heterogeneity is quantified by the calibration parameter a ($a = 1$ and $a = 0$ correspond to the uncalibrated and the fully calibrated system, respectively; for details, see Section II E). For an unconnected population of neurons subject to constant input, the width of the firing-rate distribution increases monotonically with a .

As shown in Figure 7, the level of heterogeneity (i.e., the calibration state a) is clearly reflected in the activity of the recurrent network (FB case). Both the width of the distribution of mean free membrane potentials (Figure 7a–c) as well as the width of the firing-rate distribution increases with a (Figure 7d–f; bottom panels). In the uncalibrated system ($a = 1$), a substantial fraction of neurons is predominantly driven by constant supra-threshold input currents and therefore generates highly regular spike trains ($C_V^{ISI} \approx 0$) with high firing rates ($r > 120 \text{ s}^{-1}$). Simultaneously, about 40% of the neurons are silent ($r = 0 \text{ s}^{-1}$). Neurons with intermediate firing rates ($0 \text{ s}^{-1} < r < 20 \text{ s}^{-1}$), however, show quite irregular activity ($C_V^{ISI} > 0.5$). After calibration, the firing-rate distribution is narrowed. For $a = 0$, the fraction of silent

neurons is reduced to about 10%. Maximum rates are limited to $< 80 \text{ s}^{-1}$. Note that our calibration routine compensates only for the distribution of neuron parameters, but not for the heterogeneity in synapse properties (synaptic weights, synaptic time constants; see Section IV). For the fully calibrated network ($a = 0$), the firing-rate distribution is therefore still broad. In the RAND case, we obtain similar firing-rate and inter-spike interval statistics as in the FB case (Supplements 6).

For all levels of heterogeneity attainable by our calibration procedure ($a \in [0, 1]$), input and output correlations are significantly suppressed by the recurrent-network dynamics (cf. black and dark gray vs. light gray symbols in Figure 8). In a homogeneous, random (Erdős-Rényi) network with fixed in-degree K and linear sub-threshold dynamics, the contribution of shared input to the input (free-membrane-potential) correlation is given by the network connectivity K/N [Equation 2; 5] (light gray curves in Figure 6a and Figure 8a). Nonlinearities in synaptic and/or spike-generation dynamics [89] as well as heterogeneity in neuron (and synapse) parameters lead to a suppression of this contribution [65]. Here, we refer to

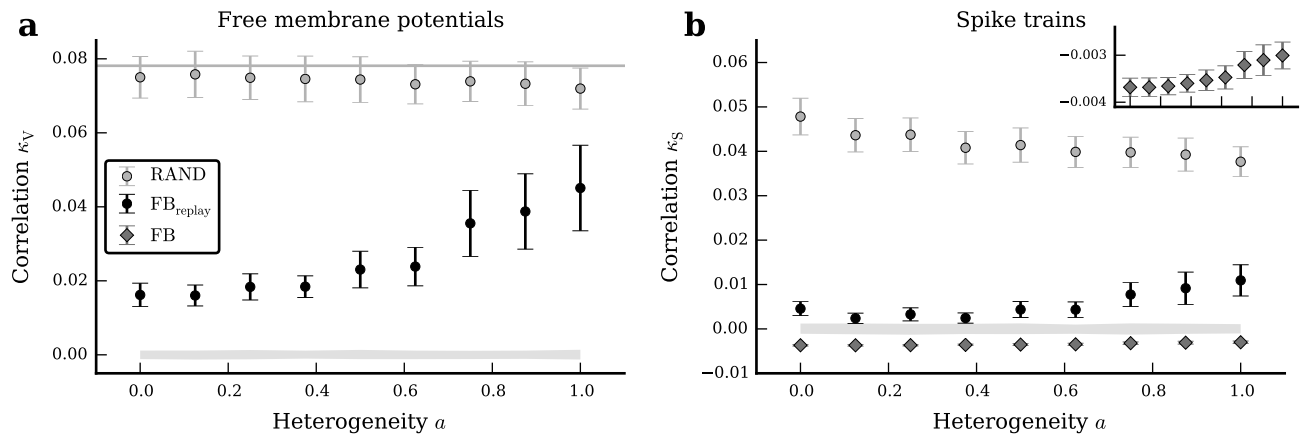


FIG. 8. Dependence of population-averaged input correlations (a) and spike-train correlations (b) on the heterogeneity of the neuromorphic substrate for the intact network (FB, dark gray diamonds), the FB_{replay} (black circles) and the RAND (light gray circles) case. Symbols and error bars denote mean and one standard deviation, respectively, across $M = 100$ network realizations (error bars are partly covered by markers). Gray curve in (a) depicts shared-input correlations in a homogeneous network (Equation 2). The inset in (b) shows a magnified view of the spike-train correlations in the FB case (dark gray diamonds). The light gray horizontal band represents mean \pm three standard deviations of correlations in surrogate data, in which spatial correlations were removed. Note that free membrane potentials cannot be recorded in the FB case (see Section II). Hence, there are no gray diamonds in (a).

this type of decorrelation as *feedforward decorrelation*. In fact, in our setup the spike-train correlations in the RAND case slightly decrease with increasing heterogeneity (light gray symbols in Figure 8b). The input correlations in the RAND case, in contrast, are marginally affected by the calibration and only slightly smaller than the theoretical value K/N (gray symbols vs. gray curve in Figure 8a). This observation may indicate that the dominant source of heterogeneity in our networks affecting correlations lies in the distributions of parameters, which affect the spike-generation (spike thresholds Θ , leak conductances g_l , resting potentials E_l) or after-spike dynamics (reset potentials v_{reset} , refractory periods τ_{ref}), but not the integration of synaptic inputs. Broad distributions of synaptic weights J , inhibitory reversal potentials E_{inh} , membrane or synaptic time constants τ_m , τ_{syn} or delays d would lead to a feedforward decorrelation also at the level of the free membrane potential. We mimicked the effect of threshold heterogeneity in network simulations on conventional computers and obtain results which are qualitatively similar to those of hardware emulations (compare Figure 9 to 8). In contrast, if we use a broad distribution of synaptic weights, we will also observe a significant decrease of input correlations (Supplements Figure 4).

Although feedforward decorrelation benefits from cell heterogeneity, input and output correlations grow with the level of heterogeneity in the presence of an intact feedback signal (black and dark gray symbols in Figure 8). We attribute this effect to a weakening of the effective feedback loop in the recurrent circuit: In heterogeneous networks with broad firing-rate distributions, neurons firing with low or high rates, corresponding to mean inputs

far below or far above firing threshold (see Figure 7a-c), are less sensitive to input fluctuations than moderately active neurons (see Supplements Figure 2). Hence, they contribute less to the overall feedback. In consequence, feedback decorrelation is impaired (see also Section IV).

IV. DISCUSSION

We have shown that inhibitory feedback effectively suppresses correlations in heterogeneous recurrent neural networks of leaky integrate-and-fire (LIF) neurons with nonlinear subthreshold dynamics, emulated on analog neuromorphic hardware (*Spikekey*; [17, 73]). Both input and output correlations are substantially smaller in networks with intact feedback loop (FB), as compared to the case where the feedback is replaced by randomized input (RAND) while preserving the connectivity structure and presynaptic firing rates. Our results hence show that active decorrelation of network activity by inhibitory feedback [4, 5] is a general phenomenon which can be observed in realistic, highly heterogeneous networks with nonlinear interaction and sufficiently strong negative feedback. Moreover, the study serves as a proof-of-principle that network activity can be efficiently decorrelated even on heterogeneous hardware, which can be exploited in functional applications, e.g., in the neuromorphic algorithms developed by Pfeil et al. [17] and Schmuker et al. [18].

Partial calibration of hardware neurons allowed us to modulate the level of network heterogeneity and, therefore, to systematically study its effect on correlations in the network activity. The analysis revealed two counter-

acting contributions: As shown in previous studies [e.g., 65], neuron heterogeneity decorrelates (shared) feedforward input (feedforward decorrelation). On the other hand, however, heterogeneity impairs feedback decorrelation (see next paragraph). In our network model, this weakening of feedback decorrelation is the dominating factor. Overall, we observed an increase in correlations with increasing level of heterogeneity. We cannot exclude that feedforward decorrelation may play a more significant role for different network configurations (e.g., different connection strengths or network topologies, different structure of external inputs, different types of heterogeneity). Our study demonstrates, however, that heterogeneity is not necessarily suppressing correlations in recurrent systems. In this context, it would be interesting to investigate the interplay of signal and noise correlations in the presence of network heterogeneities in recurrent systems. We leave this intriguing topic to future studies.

As shown in [5], feedback decorrelation in recurrent networks becomes more (less) efficient with increasing (decreasing) strength of the *effective* negative feedback. For networks of spiking neurons, the effective connection strength w_{ij} (also termed DC susceptibility [89]) between two neurons j and i corresponds to the total number of extra spikes emitted by neuron i in response to an additional input spike generated by neuron j . Assuming that the effect of a single additional input spike is small, the effective connectivity can be obtained by linear-response theory. Note that the effective weights w_{ij} depend on the working point, i.e., the average firing rates of all pre- and postsynaptic neurons (mathematically, w_{ij} is given by the derivative of the stationary response firing rate $r_i = \phi_i(r_1, \dots, r_j, \dots, r_N)$ with respect to the input firing rate r_j , evaluated at the working point; for details, see [5]). Neurons firing at very low or very high rates are typically less sensitive to input fluctuations than neurons firing at intermediate rates (due to the shape of the response function $\phi_i(r_1, \dots, r_N)$). Their dynamical range is reduced. In consequence, they hardly mediate feedback in a recurrent network. In heterogeneous networks with broad distributions of firing rates, the number of these insensitive neurons is increased. Hence, the effective feedback is weakened (see Supplements 3). We mimic the effect of heterogeneity in a linear rate model by decreasing the effective weights with increasing heterogeneity (see Supplements 5). The resulting dependence of input and output correlations on the level of heterogeneity qualitatively resembles the results we obtained for the nonlinear spiking network emulated on the neuromorphic system. A direct quantitative comparison between both models requires an explicit mapping of the synaptic weights in the LIF-neuron network to the effective weights of the linear model in the presence of distributed firing rates. We commit this task to future studies. Note that the rate dependence of the effective weights and the resulting effects of heterogeneity are consistent with our observation that LIF-neuron pairs with very low firing rates exhibit

spike-train correlations close to zero, whereas pairs with high firing rates are positively correlated. Pairs with at least one neuron firing at an intermediate rate (the second neuron can fire at a higher rate) exhibit negative spike-train correlations (see Supplements 7). As shown in [4, 5], these negative spike-train correlations are essential for compensating the positive contribution of shared inputs to the total input correlation (at least in purely inhibitory networks). Narrowing the firing rate distribution (e.g., by calibration of hardware neurons) increases the number of neurons contributing to the negative feedback, which, in turn, leads to more neuron pairs with negative spike-train correlations and, therefore, to smaller overall correlations.

Seemingly contrary to our findings, Bernacchia & Wang [72] report a decrease in correlations with increasing level of heterogeneity. The results of their study are obtained for a linear network model, which can be considered the outcome of the linearization procedure described above. Hence, the connectivity of their model corresponds to an effective connectivity (see above). Their study neglects the rate (working-point) dependence of the effective weights and can therefore not account for the effect of firing-rate heterogeneity. In [72], heterogeneity is quantified by the variance of the (effective) weight matrix (Equations 2.2 and 2.4 in [72]). For sparse connectivity matrices (with a large number of zero elements), the variance of the weight matrix reflects not only the width of the non-zero-weight distribution, but also its mean (Equation 2.4 in [72]). For networks of nonlinear spiking neurons, heterogeneities in neuron and/or synapse parameters broadens the distribution of non-zero effective weights, but may simultaneously reduce its mean (see above, Supplements 3, and [44, 95]). Hence, the variance of the full weight matrix may decrease (for illustration, see Supplements Figure 9). In other words, increasing heterogeneity in the nonlinear system may correspond to decreasing heterogeneity in the linearized system. A direct test of this hypothesis requires an explicit linearization of the nonlinear heterogeneous system.

The results of this study were obtained by network emulations on analog neuromorphic hardware. We reproduced the main findings by means of simulations of LIF-neuron networks with distributed firing thresholds on conventional computers (see Figure 9 and Supplements Figure 3). Although networks simulated on conventional computers and those emulated on the neuromorphic hardware differ in several respects (e.g., in the exact implementation of heterogeneity or the synapse model; see Supplements Table I and II), the qualitative results are very similar: In networks with intact feedback loop, input and output correlations are substantially reduced (as compared to the case where the feedback is replaced by randomized input), but increase with the extent of heterogeneity. As predicted by the theory for homogeneous inhibitory networks, we observe positive input correlations and negative output correlations (see Equation 21 in [5] and in the paragraph which fol-

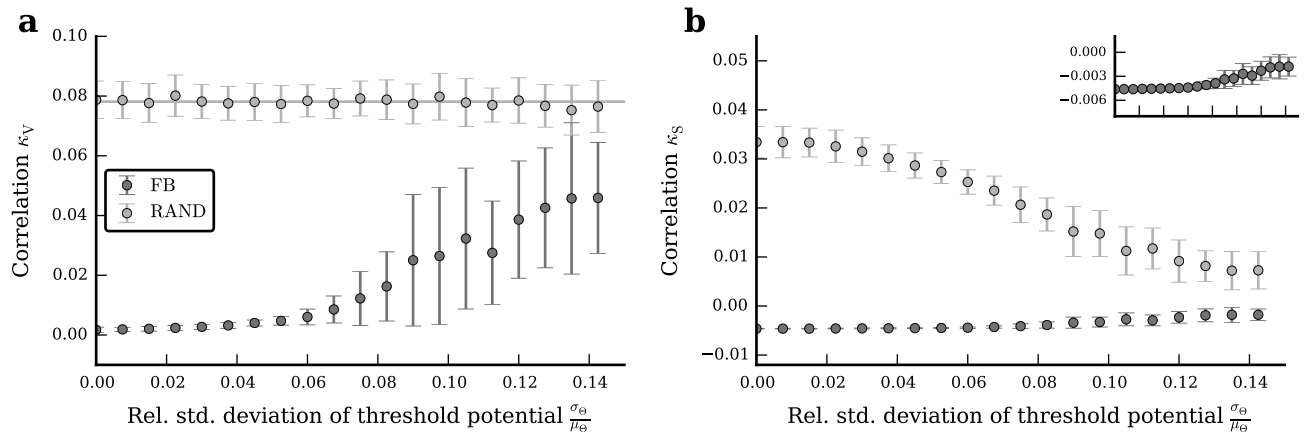


FIG. 9. Dependence of population-averaged input correlations (a), and spike-train correlations (b) on the width of threshold distributions in networks simulated with NEST [96] and PyNN [97], for the intact network (FB, dark gray circles) and the RAND (light gray circles) case. Symbols and error bars denote mean and standard deviation, respectively, across $M = 30$ network realizations (error bars are partly covered by markers). Gray curve in (a) depicts shared-input correlations in a homogeneous network (Equation 2). The inset in (b) shows a magnified view of the spike-train correlations in the FB case (dark gray circles). Note that in simulations the $\text{FB}_{\text{replay}}$ is identical to the FB case, and is hence not shown. For details see Supplements 4.

lows; see also [98] and Supplements 5). Further, note that heterogeneity in neuron parameters does not “average out” in larger networks. Upscaling the network size by a factor of 25 ($N = 4800$, in-degree $K = 375$) yields smaller spike-train correlations, but the qualitative results are similar to those obtained for the smaller network ($N = 192$, $K = 15$) emulated on the *Spikey* chip (compare Figure 8 to Supplements Figure 3).

In networks with intact feedback loop (FB scenarios), the precise spatio-temporal structure of spike trains arranges such that the self-consistent input and output correlations are suppressed. Perturbations of this structure in the local input typically lead to an increase in correlations [5]. In this study, we demonstrate this by replaying spiking activity after randomization of spike times, i.e., by replacing the time of each input spike by a random number uniformly drawn from the full emulation time interval $[0, T)$ (RAND case). However, even subtle modifications of input spike trains, such as random dither of spike times by few milliseconds, lead to an increase of correlations. On the neuromorphic hardware, replay of spike trains is not entirely reproducible (see Section II D). Hence, spike-train correlations measured in the $\text{FB}_{\text{replay}}$ mode are slightly larger than in the FB case. We would expect the same effect on the input side (free membrane potentials). Due to hardware limitations, however, we can measure input correlations only in replay mode ($\text{FB}_{\text{replay}}$ or RAND), but not in the fully connected network (FB). Therefore, all reported input correlations are likely to be slightly overestimated. In conventional network simulations, we mimicked the effect of unreliable replay by input-spike dithering and, indeed, find a gradual increase in input and output correlations (see Supplements Figure 6). These results seem to be contrary to the study by Rosenbaum & Josic [99], in which synap-

tic noise leads to a decrease of output correlations in a feedforward scenario. In our case, spike-train correlations, which suppress shared-input correlations, are removed by dithering spikes, increasing correlations on the output side. In their case, however, spike-train correlations are always zero, and shared-input correlations are decreased by synaptic failure, explaining the decreased output correlations. We attribute this contradiction to the missing feedback loop in their system, and expect correlations to increase in recurrent networks subject to similar perturbations.

Despite the imperfect replay of input spikes, the decorrelation effect is clearly visible in hardware emulations, both on the input and on the output side. The reproducibility of emulations on neuromorphic hardware could be improved by stabilizing the environment of the system, e.g., the chip temperature or the support electronics (under development). Analog hardware, however, will never reach the level of reproducibility of digital computers. But note that, similar to analog hardware, biological neurons exhibit a considerable amount of trial-to-trial variability, even under controlled in-vitro conditions [88]. So far, the details of how neuronal noise, for example, stochastic synapses (spontaneous postsynaptic events, stochastic spike transmission, synaptic failure [100]), affects correlations in recurrent neural circuits remain unclear.

Although different *Spikey* chips exhibit different realizations of fixed-pattern noise, they show a comparable extent of heterogeneity and yield results which are qualitatively similar to those presented in this article (Supplements 8). In the uncalibrated state, correlations are more sensitive to the specific realization of fixed-pattern noise and therefore vary more strongly across different chips (see (B) in Supplements Figure 13 and 14). For the

same reason, the variance of correlations across network realizations is largest in the uncalibrated state. Note that the variance across different network realizations is larger than the variance across different trials, i.e., consecutive emulations of identical networks (compare Figure 8 to Supplements Figure 12).

We have shown that negative feedback in recurrent circuits can efficiently suppress correlations, even in highly heterogeneous systems such as the analog neuromorphic architecture *Spikey*. Correlations can be further reduced by minimizing the level of network heterogeneity. In this study, we reduced the level of heterogeneity through calibration of neuron parameters in the unconnected case (see Section II E). The calibration could, in principle, be improved by calibrating neuron (and possibly synapse) parameters in the full recurrent network. Such calibration procedures are however time consuming and cumbersome. In biological substrates, homeostasis mechanisms [54, 101] keep neurons in a responsive regime and reduce the level of firing-rate heterogeneity in a self-regulating manner. Future neuromorphic devices could mimic this behavior, thereby reducing the necessity of time consuming calibration procedures. Alternatively, the analog circuits could be optimized for small parameter variations, for which likely more chip resources have to be allocated reducing the network size per chip area.

For simplicity, this work focuses on purely inhibitory networks. This demonstrates that decorrelation by inhibitory feedback does not rely on a dynamical balance between excitation and inhibition (note that the external “excitatory” drive is constant in our model) [5, 6]. Previous studies have shown that, for the homogeneous case, decorrelation by inhibitory feedback is a general phenomenon, which also occurs in excitatory-inhibitory networks, provided the overall inhibition is sufficiently strong (which is typically the case to ensure stability) [4–6, 72]. For the heterogeneous case, network simulations of excitatory-inhibitory networks show qualitatively the same results as purely inhibitory networks (compare Figure 8 to Supplements Figure 5), confirming that our results generalize to the case of mixed excitatory-inhibitory coupling.

This study demonstrates that the *Spikey* system has matured to a level that permits its use as a tool for neuroscientific research. For the results presented in this study, we recorded in total 10^{11} membrane-potential and spike-train samples, representing more than 100 days of biological time. Due to the 10^4 -fold acceleration of the *Spikey* chip, this corresponds to less than 15 minutes in the hardware-time domain. Interfacing the hardware system, however, reduces the acceleration to an approximately 50-fold speed-up (Figure 10). The translation between the network description and its hardware representation claims the majority of execution time, more than the network emulation and the transfer of data to and from the hardware system together. Encoding and decoding spike times on the host computer is especially expensive to compute. Obviously, the system could be

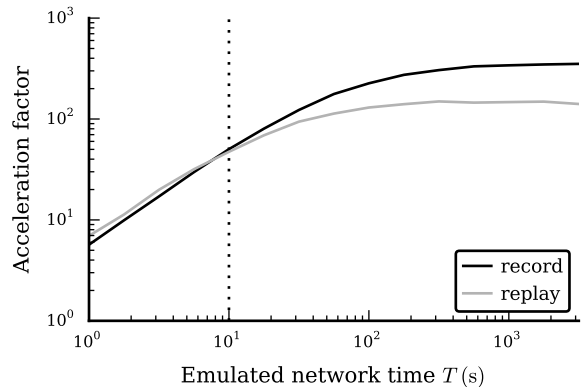


FIG. 10. Acceleration factor as a function of emulated network time T for the record (black) and the replay case (gray). The acceleration factor is defined as the ratio between the emulated network time T (in biological time) and the execution time (wall clock time). In the record case, a network realization is generated on the host computer and uploaded to the chip. During the subsequent emulation, spike trains are recorded. In the replay case, spikes are replayed and the membrane potential of one neuron is recorded with full sampling frequency (9.6 kHz). The execution time covers the full data flow from a network description in PyNN to the emulation on the *Spikey* system and back to the network representation in PyNN. The time-averaged population firing rate is $\bar{r} = (25.0 \pm 0.4) \text{ s}^{-1}$. The vertical dashed line depicts the runtime used in this study. The hardware system has to be initialized once before usage ($< 1 \text{ s}$), which is not considered here.

optimized by processing the data directly on the hardware, or by choosing a data representation which is closer to the format used on the *Spikey* chip, but this would impair user-friendliness, and hence, the effectiveness of prototyping. While the *Spikey* system permits the monitoring of the spiking activity of all neurons simultaneously, access to the membrane potentials is limited to a single (albeit arbitrary) neuron in each emulation run. Monitoring of membrane potentials of a population of n neurons therefore requires n repetitions of the same emulation. Extending the hardware system to enable access to the membrane potentials of at least two neurons simultaneously would allow for a direct observation of input correlations in the intact network (and thereby avoid problems with replay reproducibility; see above) and reduce execution time (the *Spikey* chip itself permits recording of up to eight neurons in parallel, the support electronics, however, does not). While the *Spikey* system does not significantly outperform conventional computers in terms of computational power, emulations on this system are much more energy efficient (Supplements 1). A substantial increase of computational power is expected for large systems exploiting the scalability of this technology without slow-down [102].

Functional neural architectures often rely on a stochastic dynamics of its constituents or on some form of background noise (see, e.g., [17, 18, 103]). Determinis-

tic recurrent neural networks with inhibitory feedback could provide decorrelated noise to such functional networks, both in artificial as well as in biological substrates. In neuromorphic hardware applications, these “noise networks” could thereby replace conventional random-number generators and avoid a costly transmission of background noise from a host computer to the hardware substrate (which may be particularly relevant for mobile applications with low power consumption; see Supplements 1). It needs to be investigated, however, how well functional stochastic circuits perform in the presence of such network-generated noise.

ACKNOWLEDGMENTS

We would like to thank Matthias Hock, Andreas Hartel, Eric Müller and Christoph Koke for their support

in developing and building the *Spikey* system, as well as Mihai Petrovici and Sebastian Schmitt for fruitful discussions. This research is partially supported by the Helmholtz Association portfolio theme SMHB, the Jülich Aachen Research Alliance (JARA), EU Grant 269921 (BrainScaleS), and EU Grant 604102 (Human Brain Project, HBP).

Appendix A: Network description

See Table I and II.

Appendix B: Description of data analysis

See Table III and IV.

-
- [1] Gierer, A., & Meinhardt, H. (1974). Biological pattern formation involving lateral inhibition. *Lectures on mathematics in the life sciences* 7, 163.
 - [2] Rosenzweig, M. L. (1978). Competitive speciation. *Biological Journal of the Linnean Society* 10(3), 275–289.
 - [3] Neel, L. (1952). Antiferromagnetism and ferrimagnetism. *Proceedings of the Physical Society. Section A* 65(11), 869.
 - [4] Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590.
 - [5] Tetzlaff, T., Helias, M., Einevoll, G., & Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput. Biol.* 8(8), e1002596.
 - [6] Helias, M., Tetzlaff, T., & Diesmann, M. (2014). The correlation structure of local cortical networks intrinsically results from recurrent dynamics. *PLoS Comput. Biol.* 10(1), e1003428.
 - [7] Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6* 2(11), 559–572.
 - [8] Ginis, G., & Peng, C.-N. (2006). Alien crosstalk cancellation for multipair digital subscriber line systems. *EURASIP Journal on Advances in Signal Processing* 2006(1), 016828.
 - [9] Benesty, J., Gansler, T., Morgan, D. R., Sondhi, M. M., & Gay, S. L. (2001). *Advances in Network and Acoustic Echo Cancellation* (1 ed.). Berlin, Germany: Springer.
 - [10] Trichina, E., Bucci, M., De Seta, D., & Luzzi, R. (2001). Supplemental cryptographic hardware for smart cards. *Micro, IEEE* 21(6), 26–35.
 - [11] Mead, C. (1990). Neuromorphic electronic systems. *Proc. IEEE* 78(10), 1629–1636.
 - [12] Indiveri, G., Linares-Barranco, B., Hamilton, T. J., van Schaik, A., Etienne-Cummings, R., Delbruck, T., Liu, S.-C., Dudek, P., Häfliger, P., Renaud, S., Schemmel, J., Cauwenberghs, G., Arthur, J., Hynna, K., Folowosele, F., Saighi, S., Serrano-Gotarredona, T., Wijekoon, J., Wang, Y., & Boahen, K. (2011). Neuromorphic silicon neuron circuits. *Front. Neurosci.* 5(73).
 - [13] SenseMaker (2009). SenseMaker: A multi-sensory, task-specific, adaptable perception system; project website. Available at: http://cordis.europa.eu/project/rcn/62746_en.html.
 - [14] FACETS (2010). Fast Analog Computing with Emergent Transient States, project website. Available at: <http://www.facets-project.org>.
 - [15] BrainScaleS (2014). Project website. Available at: <http://www.brainscales.eu>.
 - [16] Human Brain Project (2014). Project website. Available at: <http://www.humanbrainproject.eu>.
 - [17] Pfeil, T., Grübl, A., Jeltsch, S., Müller, E., Müller, P., Petrovici, M. A., Schmuker, M., Brüderle, D., Schemmel, J., & Meier, K. (2013). Six networks on a universal neuromorphic computing substrate. *Front. Neurosci.* 7(11).
 - [18] Schmuker, M., Pfeil, T., & Nawrot, M. P. (2014). A neuromorphic network for generic multivariate data classification. *Proc. Natl. Acad. Sci. USA* 111(6), 2081–2086.
 - [19] Ecker, A. S., Berens, P., Tolia, A. S., & Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.* 31(40), 14272–14283.
 - [20] Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366.
 - [21] Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417.
 - [22] Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9(10), 474–480.
 - [23] Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex* (1st ed.). Cambridge: Cambridge University Press.
 - [24] Diesmann, M., Gewaltig, M.-O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402(6761), 529–533.

A			Model summary																				
Populations			One (inhibitory)																				
Topology			-																				
Connectivity			Random convergent connections (fixed in-degree)																				
Neuron model			Leaky integrate-and-fire (LIF), fixed firing threshold, fixed absolute refractory time																				
Channel models			-																				
Synapse model			Exponentially decaying conductances, fixed delays																				
Plasticity			-																				
External input			Resting potential higher than threshold (= constant current) ($E_l > \Theta$)																				
Measurements			Spikes and membrane potentials from all neurons																				
Other			No autapses, no multapses																				
B						Populations																	
Name			Elements			Size																	
I			LIF neuron			N																	
C												Connectivity											
Source		Target		Pattern																			
I		I		Random convergent connect, in-degree K																			
D																		Neuron and synapse model					
Type						Leaky integrate-and-fire, exponential conductances																	
Subthreshold dynamics						Subthreshold dynamics ($t \notin (t^*, t^* + \tau_{\text{ref}})$): $C_m \frac{d}{dt} v(t) = -g_l(v(t) - E_l) - g_{\text{syn}}(t)(v(t) - E_{\text{inh}})$ Reset and refractoriness ($t \in (t^*, t^* + \tau_{\text{ref}})$): $v(t) = v_{\text{reset}}$ This model is emulated by analog circuitry on the <i>Spikey</i> chip [12].																	
Conductance dynamics						For each presynaptic spike at time t^* ($t > t^* + d$): $g_{\text{syn}}(t) \approx J \exp(-\frac{t-t^*-d}{\tau_{\text{syn}}})$, where $J = w_{\text{hw}} g_{\text{max}}$ This model is emulated by analog circuitry on the <i>Spikey</i> chip [17].																	
Spiking						If $v(t^*-) < \Theta \wedge v(t^*+) \geq \Theta$: emit spike with time stamp t^*																	

TABLE I. Description of the network model (according to [104]).

- [25] Tabareau, N., Slotine, J.-J., & Pham, Q.-C. (2010). How synchronization protects from noise. *PLoS Comput. Biol.* 6(1). doi:10.1371/journal.pcbi.1000637.
- [26] Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2(8), 539–550.
- [27] Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.* 18(10), 3870–3896.
- [28] Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11, 91–101.
- [29] Tripp, B., & Eliasmith, C. (2007). Neural populations can induce reliable postsynaptic currents without observable spike rate changes or precise spike timing. *Cereb. Cortex* 17(8), 1830–1840.
- [30] Schmuker, M., & Schneider, G. (2007). Processing and classification of chemical data inspired by insect olfaction. *Proc. Natl. Acad. Sci. USA* 104(51), 20285–20289.
- [31] Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600.
- [32] Tetzlaff, T., Morrison, A., Geisel, T., & Diesmann, M. (2004). Consequences of realistic network size on the stability of embedded synfire chains. *Neurocomputing* 58–60, 117–121.

B Populations		
Name	Values	Description
N	{96, 112, 128, 144, 160, 176, 192 }	network size
C Connectivity		
Name	Values	Description
K	15	number of presynaptic partners
D Neuron		
Name	Values	Description
C_m	0.2 nF	membrane capacitance
τ_{ref}	1 ms	refractory period
v_{reset}	−80 mV	reset potential
E_l	−52 mV	resting potential
Θ	−62 mV	firing threshold
g_l	calibrated	leak conductance
E_{inh}	−80 mV	inhibitory reversal potential
D Synapse		
Name	Values	Description
g_{max}	in the order of 1 nS	conductance amplitude
τ_{syn}	$\ll \frac{C_m}{g_l}$	conductance time constant
w_{hw}	3	synaptic weight (in hardware values $\in [0, 15]$)
d	≈ 1.3 ms	synaptic delay
Other Software		
Name	Values	Description
	67bc2eec	git revision SpikeyHAL
	e793bb97	git revision pyNN
	d948716a	git revision vmodule

TABLE II. Parameter values for the network model described in Table I. Bold numbers indicate default values. Leak conductances g_l are adjusted in the calibration process (see Section II E). Gray numbers indicate target values not considering fixed-pattern noise.

- [33] Kriener, B., Tetzlaff, T., Aertsen, A., Diesmann, M., & Rotter, S. (2008). Correlations and population dynamics in cortical networks. *Neural Comput.* 20, 2185–2226.
- [34] Tetzlaff, T., Rotter, S., Stark, E., Abeles, M., Aertsen, A., & Diesmann, M. (2008). Dependence of neuronal correlations on filter characteristics and marginal spike-train statistics. *Neural Comput.* 20(9), 2133–2184.
- [35] Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., & Logothetis, N. K. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327(5965), 584–587.
- [36] Ly, C., Middleton, J., & Doiron, B. (2012). Cellular and circuit mechanisms maintain low spike co-variability and enhance population coding in somatosensory cortex. *Front. Comput. Neurosci.* 6(7), 1–25.
- [37] Middleton, J., Omar, C., Doiron, B., & Simons, D. (2012). Neural correlation is stimulus modulated by feedforward inhibitory circuitry. *J. Neurosci.* 32(2), 506–518.
- [38] Wiechert, M. T., Judkewitz, B., Riecke, H., & Friedrich, R. W. (2010). Mechanisms of pattern decorrelation by recurrent neuronal circuits. *Nat. Neurosci.* 13, 1003–1010.
- [39] Griffith, J. S., & Horn, G. (1966). An analysis of spontaneous impulse activity of units in the striate cortex of unrestrained cats. *J. Physiol. (Lond.)* 186(3), 516–534.
- [40] Koch, K. W., & Fuster, J. M. (1989). Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Exp. Brain Res.* 76(2), 292–306.

A Analysis measures	
Measure	Details
spike density	$\xi_i(t) = \sum_k \delta(t - t_i^k)$
spike train	$s_i(t_k) = \text{number of spikes of neuron } i \text{ per bin } [k\Delta t, (k+1)\Delta t)$
population activity	$\bar{s}(t) = \frac{1}{N} \sum_i s_i(t)$
time-averaged population activity	$\bar{r} = \langle \bar{s}(t) \rangle_t$
membrane potential	$v_i(t_k) = \text{membrane potential of neuron } i \text{ at time step } k$
(finite time) Fourier transform	$X_i(f) = \mathfrak{F}[x_i(t)](f) = \int_0^T dt x_i(t) e^{-2\pi i f t}$ (with inverse \mathfrak{F}^{-1})
(single unit) power spectrum	$A_i(f) = \frac{1}{T} X_i^*(f) X_i(f)$
population-averaged power spectrum	$A(f) = \frac{1}{N} \sum_i A_i(f)$
population power spectrum	$\bar{A}(f) = (\sum_i S_i^*(f))(\sum_j S_j(f))$
pairwise cross spectrum	$C_{ij} = \frac{1}{T} X_i^*(f) X_j(f), i \neq j$
population-averaged cross spectrum	$C(f) = \frac{1}{N(N-1)} \sum_{i \neq j} C_{ij}(f) \equiv \frac{1}{N(N-1)} (\bar{A}(f) - NA(f))$ (note: $C(f) \in \mathbb{R}$)
sliding window filter	$X(f) \rightarrow X(f) * H(f)$ with $H(f) = \frac{1}{f_1 - f_0} \Theta(f - f_0) \Theta(f_1 - f)$
coherence	$\kappa(f) = \frac{C(f)}{A(f)}$
low-frequency coherence	$\kappa_X = \frac{1}{f_{\max} - f_{\min}} \int_{f_{\min}}^{f_{\max}} df \kappa(f)$
pop.-averaged cross-correlation function	$c(\tau) = \frac{1}{N(N-1)} \sum_{i \neq j} \langle s_i(t) s_j(t + \tau) \rangle_t \equiv \mathfrak{F}^{-1}[C(f)](\tau)$
time average	$\langle \dots \rangle_t$

TABLE III. Summary of the data analysis. Here $i \in [1, N]$, $X \in S, V$.

A Analysis parameters		
Parameter	Description	Values
Δt	bin size for spike trains	1 ms
Δt_m	bin size for membrane potential traces	0.52 ms
T_{warmup}	initial warm-up time (not considered in analysis)	1 s
T	emulated network time (biol. time domain)	10 s
M	number of network realizations	{50, 100 }
ΔF	width of sliding window	1 Hz
f_{\min}, f_{\max}	interval boundaries for low-frequency coherence	0.1 Hz, 20 Hz
a	calibration state	$\{0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}, \mathbf{1}\}$

TABLE IV. Summary of analysis parameters (default values in bold).

- [41] Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., & Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* 146(3), 1082–1108.
- [42] Hromadka, T., DeWeese, M. R., & Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized

auditory cortex. *PLoS Biol.* 6(1).

- [43] O'Connor, D. H., Peron, S. P., Huber, D., & Svoboda, K. (2010). Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* 67(6), 1048–1061.

- [44] Roxin, A., Brunel, N., Hansel, D., Mongillo, G., & van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. *J. Neurosci.* 31(45), 16217–16226.
- [45] Buzsaki, G., & Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* 15, 264–278.
- [46] Song, S., Sjöström, P., Reigl, M., Nelson, S., & Chklovskii, D. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* 3(3), e68.
- [47] Lefort, S., Tómm, C., Sarría, J.-C. F., & Petersen, C. C. H. (2009). The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* 61, 301–316.
- [48] Koulakov, A. A., Hromádka, T., & Zador, A. M. (2009). Correlated connectivity and the distribution of firing rates in the neocortex. *J. Neurosci.* 29(12), 3685–3694.
- [49] Avermann, M., Tómm, C., Mateo, C., Gerstner, W., & Petersen, C. (2012). Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *J. Neurophysiol.* 107(11), 3116–3134.
- [50] Ikegaya, Y., Sasaki, T., Ishikawa, D., Honma, N., Tao, K., Takahashi, N., Minamisawa, G., Ujita, S., & Matsuki, N. (2013). Interpyramidal spike transmission stabilizes the sparseness of recurrent network activity. *Cereb. Cortex* 23(2), 293–304.
- [51] Roxin, A. (2011). The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons. *Front. Comput. Neurosci.* 5(8).
- [52] Kuhn, A., Aertsen, A., & Rotter, S. (2004). Neuronal integration of synaptic input in the fluctuation-driven regime. *J. Neurosci.* 24(10), 2345–2356.
- [53] Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: Noise or part of the signal? *Nat. Rev. Neurosci.* 6, 389–397.
- [54] Marder, E., & Goaillard, J.-M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nat. Rev. Neurosci.* 7(7), 563–574.
- [55] Van Vreeswijk, C., & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 10(6), 1321–1371.
- [56] Tsodyks, M., Mitkov, I., & Sompolinsky, H. (1993). Pattern of synchrony in inhomogeneous networks of oscillators with pulse interactions. *Phys. Rev. Lett.* 71(8).
- [57] Golomb, D., & Rinzel, J. (1993). Dynamics of globally coupled inhibitory neurons with heterogeneity. *Phys. Rev. E* 48(6), 4810–4814.
- [58] Neltner, L., Hansel, D., Mato, G., & Meunier, C. (2000). Synchrony in heterogeneous networks of spiking neurons. *Neural Comput.* 12(7), 1607–1641.
- [59] Denker, M., Timme, M., Diesmann, M., Wolf, F., & Geisel, T. (2004). Breaking synchrony by heterogeneity in complex networks. *Phys. Rev. Lett.* 92(7), 074103–1–074103–4.
- [60] Mejias, J. F., & Longtin, A. (2012). Optimal heterogeneity for coding in spiking neural networks. *Phys. Rev. Lett.* 108.
- [61] Stocks, N. G. (2000). Suprathreshold stochastic resonance in multilevel threshold systems. *Physical Review Letters* 84(11), 2310.
- [62] Shamir, M., & Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. *Neural computation* 18(8), 1951–1986.
- [63] Chelaru, M. I., & Dragoi, V. (2008). Efficient coding in heterogeneous neuronal populations. *Proc. Natl. Acad. Sci. USA* 105(42).
- [64] Osborne, L. C., Palmer, S. E., Lisberger, S. G., & Bialek, W. (2008). The neural basis for combinatorial coding in a cortical population response. *The Journal of Neuroscience* 28(50), 13522–13531.
- [65] Padmanabhan, K., & Urban, N. N. (2010). Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nat. Neurosci.* 13(10), 1276–1282.
- [66] Marsat, G., & Maler, L. (2010). Neural heterogeneity and efficient population codes for communication signals. *Journal of Neurophysiology* 104(5), 2543–2555.
- [67] Holmström, L. A., Eeuwes, L. B., Roberts, P. D., & Portfors, C. V. (2010). Efficient encoding of vocalizations in the auditory midbrain. *The Journal of Neuroscience* 30(3), 802–819.
- [68] Yim, M. Y., Aertsen, A., & Rotter, S. (2013). Impact of intrinsic biophysical diversity on the activity of spiking neurons. *Phys. Rev. E* 87.
- [69] Lengler, J., Jug, F., & Steger, A. (2013). Reliable neuronal systems: The importance of heterogeneity. *PLoS One*.
- [70] Mejias, J. F., & Longtin, A. (2014). Differential effects of excitatory and inhibitory heterogeneity on the gain and asynchronous state of sparse cortical networks. *Frontiers in computational neuroscience* 8.
- [71] Tripathy, S. J., Padmanabhan, K., Gerkin, R. C., & Urban, N. N. (2013). Intermediate intrinsic diversity enhances neural population coding. *Proc. Natl. Acad. Sci. USA* 110(20), 8248–8253.
- [72] Bernacchia, A., & Wang, X.-J. (2013). Decorrelation by recurrent inhibition in heterogeneous neural circuits. *Neural Comput.* 25(7), 1732–1767.
- [73] Schemmel, J., Grünbl, A., Meier, K., & Müller, E. (2006). Implementing synaptic plasticity in a VLSI spiking neural network model. In *Proceedings of the 2006 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, pp. 1–6. IEEE Press.
- [74] Badoni, D., Giulioni, M., Dante, V., & Del Giudice, P. (2006). An aVLSI recurrent network of spiking neurons with reconfigurable and plastic synapses. In *Proceedings of the 2006 International Symposium on Circuits and Systems (ISCAS)*, Island of Kos, pp. 4. IEEE Press.
- [75] Häfziger, P. (2007). Adaptive WTA with an analog VLSI neuromorphic learning chip. *IEEE Trans. Neural Netw.* 18(2), 551–572.
- [76] Vogelstein, R., Mallik, U., Vogelstein, J., & Cauwenberghs, G. (2007). Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses. *IEEE Trans. Neural Netw.* 18(1), 253–265.
- [77] Indiveri, G., Chicca, E., & Douglas, R. (2009). Artificial cognitive systems: From VLSI networks of spiking neurons to neuromorphic cognition. *Cognitive Computation* 1(2), 119–127.
- [78] Serrano-Gotarredona, R., Oster, M., Lichtsteiner, P., Linares-Barranco, A., Paz-Vicente, R., Gomez-Rodriguez, F., Camunas-Mesa, L., Berner, R., Rivas-Perez, M., Delbruck, T., Liu, S.-C., Douglas, R., Häfziger, P., Jimenez-Moreno, G., Ballcells, A., Serrano-Gotarredona, T., Acosta-Jimenez, A., & Linares-Barranco, B. (2009). CAVIAR: A 45k neuron, 5M synapse, 12G connects/s AER hardware sensory - pro-

- cessing - learning - actuating system for high-speed visual object recognition and tracking. *IEEE Trans. Neural Netw.* 20(9), 1417–1438.
- [79] Brink, S., Nease, S., Hasler, P., Ramakrishnan, S., Wunderlich, R., Basu, A., & Degnan, B. (2013). A learning-enabled neuron array IC based upon transistor channel models of biological phenomena. *IEEE Trans. Biomed. Circuits Syst.* 7(1), 71–81.
- [80] Benjamin, B., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A., Bussat, J., Alvarez-Icaza, R., Arthur, J., Merolla, P., & Boahen, K. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proc. IEEE* 102(5), 699–716.
- [81] Renaud, S., Tomas, J., Lewis, N., Bornat, Y., Daouzli, A., Rudolph, M., Destexhe, A., & Saïghi, S. (2010). PAX: A mixed hardware/software simulation platform for spiking neural networks. *Neural Networks* 23(7), 905–916.
- [82] Yu, T., & Cauwenberghs, G. (2010). Analog VLSI biophysical neurons and synapses with programmable membrane channel kinetics. *IEEE Trans. Biomed. Circuits Syst.* 4(3), 139–148.
- [83] Davison, A., Brüderle, D., Eppler, J. M., Kremkow, J., Müller, E., Pecevski, D., Perrinet, L., & Yger, P. (2009). PyNN: a common interface for neuronal network simulators. *Front. Neuroinformatics* 2(11).
- [84] Brüderle, D., Müller, E., Davison, A., Müller, E., Schemmel, J., & Meier, K. (2009). Establishing a novel modeling tool: A Python-based interface for a neuromorphic hardware system. *Front. Neuroinformatics* 3(17).
- [85] Kaplan, B., Brüderle, D., Schemmel, J., & Meier, K. (2009). High-conductance states on a neuromorphic hardware system. In *Proceedings of the 2009 International Joint Conference on Neural Networks (IJCNN)*, Atlanta, pp. 1524–1530. IEEE Press.
- [86] Bill, J., Schuch, K., Brüderle, D., Schemmel, J., Maass, W., & Meier, K. (2010). Compensating inhomogeneities of neuromorphic VLSI devices via short-term synaptic plasticity. *Front. Comput. Neurosci.* 4(129).
- [87] Brüderle, D., Bill, J., Kaplan, B., Kremkow, J., Meier, K., Müller, E., & Schemmel, J. (2010). Simulator-like exploration of cortical network architectures with a mixed-signal VLSI system. In *Proceedings of the 2010 International Symposium on Circuits and Systems (ISCAS)*, Paris, France. IEEE Press.
- [88] Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science* 268, 1503–1506.
- [89] De la Rocha, J., Doiron, B., Shea-Brown, E., Kresimir, J., & Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448(16), 802–807.
- [90] Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). Cambridge University Press.
- [91] Pfeil, T., Scherzer, A.-C., Schemmel, J., & Meier, K. (2013). Neuromorphic learning towards nano second precision. In *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pp. 1–5. IEEE Press.
- [92] Bair, W., Zohary, E., & Newsome, W. (2001). Correlated firing in Macaque visual area MT: time scales and relationship to behavior. *J. Neurosci.* 21(5), 1676–1697.
- [93] Kohn, A., & Smith, M. A. (2005). Stimulus dependence of neuronal correlations in primary visual cortex of the Macaque. *J. Neurosci.* 25(14), 3661–3673.
- [94] Moreno-Bote, R., & Parga, N. (2006). Auto- and cross-correlograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.* 96, 028101.
- [95] Grytskyy, D., Tetzlaff, T., Diesmann, M., & Helias, M. (2013). A unified view on weakly correlated recurrent networks. *Front. Comput. Neurosci.* 7, 131.
- [96] Gewaltig, M.-O., & Diesmann, M. (2007). NEST (NEural Simulation Tool). *Scholarpedia* 2(4), 1430.
- [97] Davison, A., Brüderle, D., Eppler, J., Kremkow, J., Müller, E., Pecevski, D., Perrinet, L., & Yger, P. (2008). PyNN: a common interface for neuronal network simulators. *Front. Neuroinformatics* 2(11). doi:10.3389/neuro.11.011.2008.
- [98] Helias, M., Tetzlaff, T., & Diesmann, M. (2013). Echoes in correlated neural systems. *New J. Phys.* 15, 023002.
- [99] Rosenbaum, R., & Josic, K. (2011). Mechanisms that modulate the transfer of spiking correlations. *Neural Comput.* 23(5), 1261–1305.
- [100] Ribault, C., Sekimoto, K., & Triller, A. (2011). From the stochasticity of molecular processes to the variability of synaptic transmission. *Nat. Rev. Neurosci.* 12, 375–387.
- [101] O’Leary, T., Williams, A. H., Franci, A., & Marder, E. (2014). Cell types, network homeostasis, and pathological compensation from a biologically plausible ion channel expression model. *Neuron* 82(4), 809–821.
- [102] Schemmel, J., Brüderle, D., Grünbl, A., Hock, M., Meier, K., & Millner, S. (2010). A wafer-scale neuromorphic hardware system for large-scale neural modeling. In *Proceedings of the 2010 International Symposium on Circuits and Systems (ISCAS)*, Paris, pp. 1947–1950. IEEE Press.
- [103] Maass, W. (2014). Noise as a resource for computation and learning in networks of spiking neurons. *Proc. IEEE* 102(5), 860–880.
- [104] Nordlie, E., Gewaltig, M.-O., & Plesser, H. E. (2009). Towards reproducible descriptions of neuronal network models. *PLoS Comput. Biol.* 5(8), e1000456.
- [105] Sharp, T., Galluppi, F., Rast, A., & Furber, S. (2012). Power-efficient simulation of detailed cortical microcircuits on SpiNNaker. *J. Neurosci. Methods* 210(1), 110–118.

Supplements

Supplements 1: Power consumption

The *Spikey* system consumes approximately 6 W of power, and the chip itself less than 0.6 W. On the chip most power is consumed by digital communication infrastructure, which is not part of the neuromorphic network. In the following, we estimate the power consumption for a single synaptic event using the data set partly shown in Figure 5a. This emulation lasts $T = 10$ s in biological time and generates approximately $48 \cdot 10^3$ spikes. Considering the acceleration of the hardware network (10^4) and the synapse count per neuron ($K = 15$), the system generates $720 \cdot 10^6$ synaptic events per second in hardware time. If we consider the total power consumption of the *Spikey* chip, the upper bound of energy consumed by each synaptic transmission will be approximately 1 nJ. Because these measurements include the communication infrastructure and other support electronics to observe spike times and membrane traces, the real energy consumption for synaptic transmissions is estimated to be approximately ten times smaller. Network simulations on conventional supercomputers are far less energy efficient and consume tens of μ J for each synaptic transmission [105].

Supplements 2: Modification of the bisection method

In each iteration of the bisection method that is used to calibrate the leak conductances of hardware neurons (Section II E), we evaluated the firing rate for each neuron by the median over $L = 25$ identical trials. However, if this measure is compared between consecutive identical iterations, temporal noise on time scales longer than the duration of one iteration may still lead to variability. In the original bisection method, the interval of possible solutions is halved after each iteration step [90]. To improve the convergence of this method in the context of our calibration we expanded the halved interval by 20% at both ends after each iteration. This prevents the algorithm to get stuck in an interval wrongly chosen by random fluctuations of the firing rates.

Supplements 3: Effective weights

We quantify the effect of a single spike of neuron j on the firing rate of a postsynaptic neuron i by the effective weight w_{ij} of the connection $i \leftarrow j$. Assuming that the activity of neuron i does not affect the activity of neuron j (i.e., the RAND case), we define w_{ij} as the cross-correlation between the spike trains $s_j(t)$ and $s_i(t + \tau)$, where, due to causality, τ is positive. Then, we average the effective weight over the emulated network time T , and subtract the baseline determined by the average correlation for negative τ :

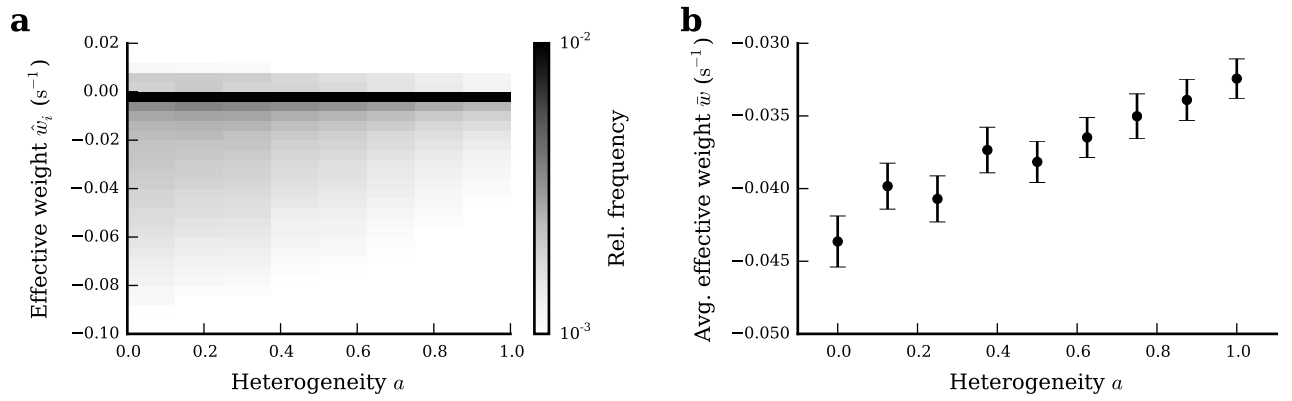
$$w_{ij} = \frac{1}{\tau_{\max}} \int_0^{\tau_{\max}} \langle s_j(t) s_i(t + \tau) \rangle_t d\tau - \frac{1}{\tau_{\min}} \int_{-\tau_{\min}}^0 \langle s_j(t) s_i(t + \tau) \rangle_t d\tau \quad . \quad (\text{S1})$$

Here, we chose $\tau_{\max} = 50$ ms and $\tau_{\min} = -50$ ms. τ_{\max} was determined by measuring the average duration in which a spike from neuron j has an influence on neuron i (data not shown). τ_{\min} was then chosen symmetrically. The density of effective weights shifts towards less negative effective weights for increasing heterogeneity (Supplements Figure 1a).

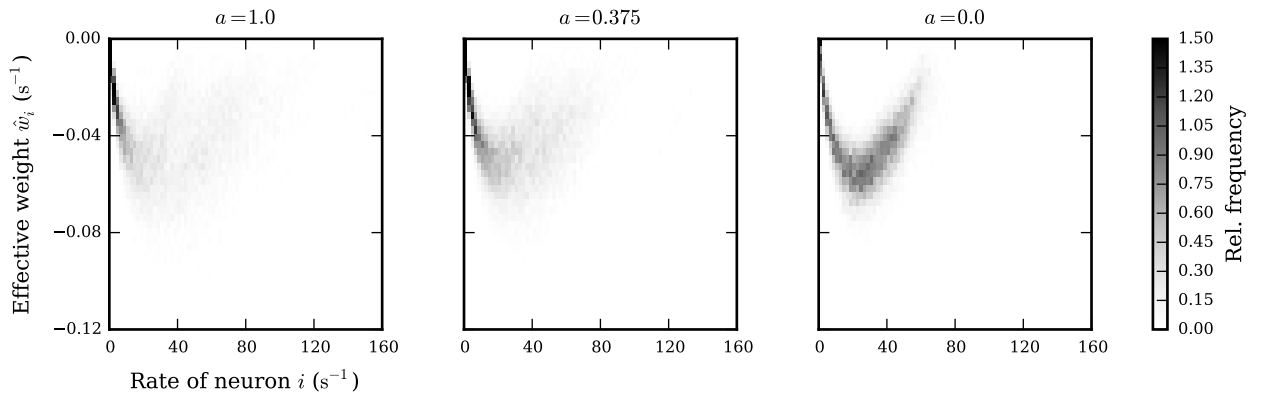
We obtain \bar{w} by averaging over all possible connections:

$$\bar{w} = \frac{1}{NK} \sum_{i,j} w_{ij} \quad . \quad (\text{S2})$$

For increasing heterogeneity the average effective weight \bar{w} is less negative (Supplements Figure 1b). This can be explained by the dependence of the effective weight on the firing rate of the postsynaptic neuron (Supplements Figure 2). In the regime of small rates ($< 10 \text{ s}^{-1}$), incoming spikes hardly affect the neuron's firing and the effective weight is small. Similarly, in the case for large rates ($> 40 \text{ s}^{-1}$). Neurons with intermediate firing rates are sensitive to input and hence have a more negative effective weight. Heterogeneity increases the number of neurons with small and large firing rates, and hence the average effective weight is less negative, which in turn weakens the negative feedback of the network.



SUP. FIG. 1. Distribution of effective weights w (a; log scale) and population-averaged effective weights \bar{w} (b) for different levels of heterogeneity, averaged over $M = 100$ network realizations. Symbols and error bars denote mean and one standard deviation, respectively.

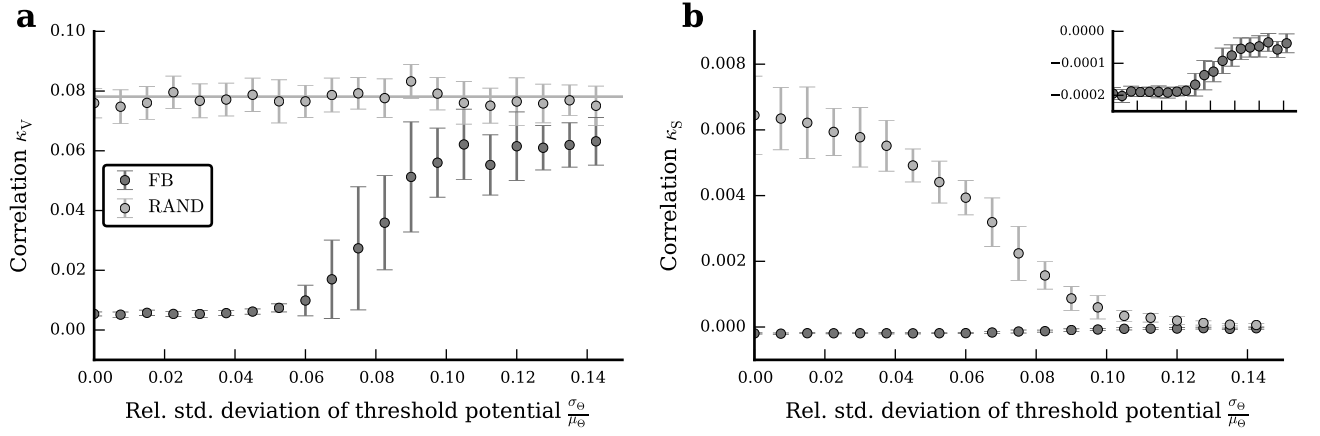


SUP. FIG. 2. Distribution of average incoming effective weights \hat{w}_i in dependence of the rate of the postsynaptic neuron i , for three different levels of heterogeneity ($a = 0$, $a = 0.375$ and $a = 1$), averaged over $M = 100$ network realizations. The average incoming weight is defined as $\hat{w}_i = \sum_j w_{ij}$.

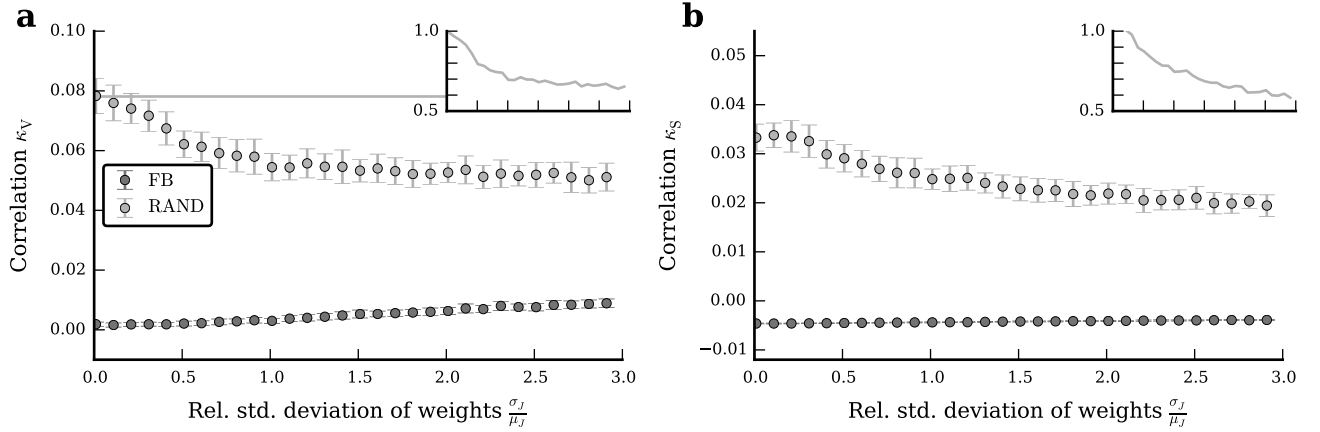
Supplements 4: Simulations with software

We validate our results by comparing them to simulations with software (NEST [96], PyNN [97]). In these simulations we modulated the degree of heterogeneity by distributing the firing thresholds of all neurons according to a normal distribution with mean Θ and variance σ_Θ . Details about the network, neuron and synapse models and their parameters can be found in Supplements Table I and II, respectively. The results are qualitatively the same compared to network emulations on the *Spikey* chip (compare Figure 8 to 9) and also hold for larger network sizes (Supplements Figure 3). In the FB case, input correlations increase with network heterogeneity and spike-train correlations become less negative. In the RAND case, input correlations stay approximately constant, while output correlations decrease with the variance σ_Θ . This can be explained by the fact that, here, heterogeneity only affects the output spike times, and not the integrative properties of the neurons (see also Section III B and [68]).

In addition, we compare the effect of heterogeneity in firing thresholds on correlations to that of heterogeneity in synaptic weights. Details about the network, neuron and synapse models and their parameters can be found in Supplements Table I and III, respectively. In the RAND case, we find that a distribution of firing thresholds reduces, as expected, only output correlations (Figure 9). If we distribute weights instead, shared-input correlations are reduced (Supplements Figure 4). Output correlations are also reduced, however, only proportionally to the reduction of input correlations (compare insets in Supplements Figure 4). While output correlations are overall smaller than input correlations due to the non-linearity in spike generation, we do not observe a boost of this decrease for large heterogeneities. Overall, the dynamics of the recurrent system is more sensitive to heterogeneities in firing thresholds than in synaptic weights (compare scale of abscissas of Figure 9 to Supplements Figure 4). For hardware emulations we measure a reduction of output correlations, but only a minor change in input correlations (see RAND case in Figure 8),



SUP. FIG. 3. Like Figure 9, but for $N = 4800$ neurons and $K = 375$ inputs for each neuron. Note the different scales of the ordinate in Figure 9 and (b). Symbols and error bars denote mean and standard deviation, respectively, across $M = 20$ network realizations (error bars are partly covered by markers).

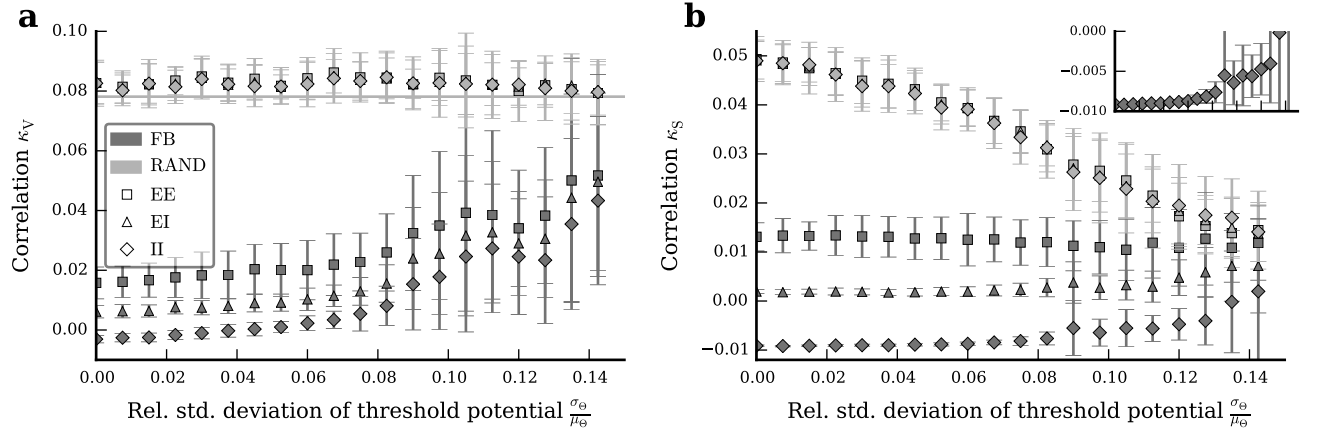


SUP. FIG. 4. Dependence of population-averaged input correlations (a), and spike-train correlations (b) on the width of the weight distribution, for the intact network (FB, dark gray circles) and the RAND (light gray circles) case. Symbols and error bars denote mean and standard deviation, respectively, across $M = 30$ network realizations (error bars are partly covered by markers). Gray curve in (a) depicts shared-input correlations in a homogeneous network (Equation 2). The inset shows correlations normalized to unity at $\sigma_J = 0$, with the same abscissa as in the main plot. Note that in simulations the $\text{FB}_{\text{replay}}$ is identical to the FB case, and is hence not shown. Networks simulated with NEST [96] and PyNN [97].

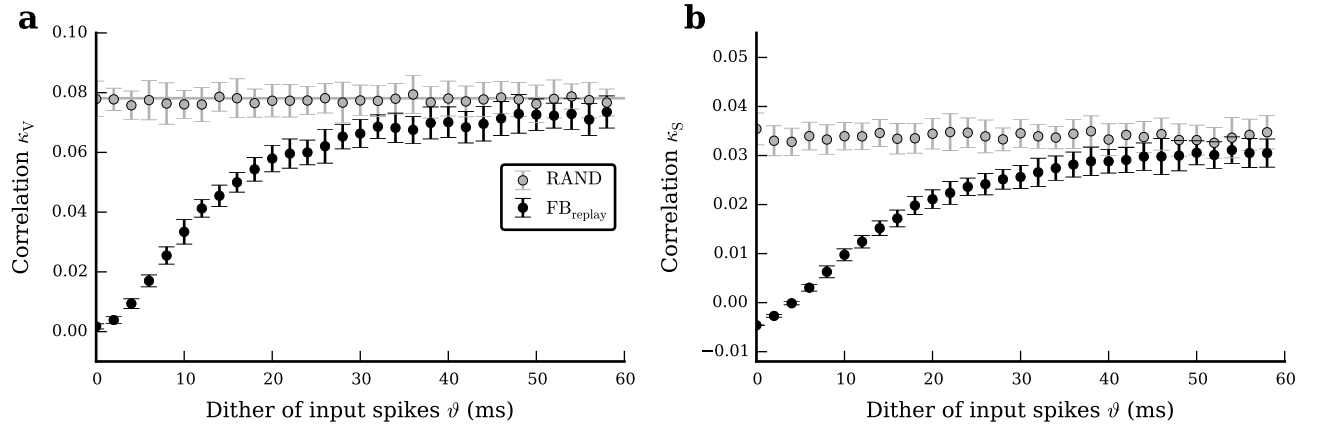
suggesting that the main source of heterogeneity affecting correlations lies in single neuron parameters governing the excitability of neurons. This suggestion is further supported by the observation that on hardware input correlations in the RAND case are only insignificantly smaller than the theoretical value for networks with homogeneous weights, although weights are distributed on hardware. In the FB case, we observe a strong sensitivity of input correlations to heterogeneity both for hardware emulations and networks with distributed thresholds, but not for networks with distributed weights.

To investigate the generalization of our results to mixed excitatory-inhibitory networks, correlations were measured in a network of $N = 192$ neurons consisting of half excitatory and half inhibitory neurons. Details about the network, neuron and synapse models and their parameters can be found in Supplements Table IV and V, respectively. As described above, we distribute the firing thresholds of all neurons according to a normal distribution. The results are consistent with those obtained from purely inhibitory networks on hardware demonstrating the generality of our findings (compare Figure 8 to Supplements Figure 5). By modulating the level of heterogeneity separately for the excitatory or the inhibitory population, we observe that network dynamics are more sensitive to heterogeneities in the inhibitory than in the excitatory population (data not shown).

We investigate the effect of temporal noise on correlations by dithering spikes in the $\text{FB}_{\text{replay}}$ case before replaying them to the network (for network, neuron and synapse models see Supplements Table I and II). Each spike time t is

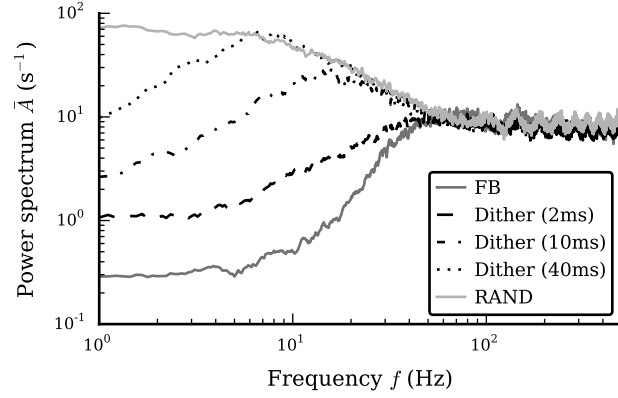


SUP. FIG. 5. Dependence of population-averaged input correlations (a), and spike-train correlations (b) on the width of the threshold distribution, for the intact network (FB, dark gray circles) and the RAND (light gray circles) case. The network consists of two populations, one excitatory and one inhibitory, of equal size. Symbols and error bars denote mean and standard deviation, respectively, across $M = 30$ network realizations (error bars are partly covered by markers). Gray curve in (a) depicts shared-input correlations in a homogeneous network (Equation 2). The inset in (b) shows a magnified view of the spike-train correlations in the FB case (dark gray diamonds). Note that in simulations the $\text{FB}_{\text{replay}}$ is identical to the FB case, and is hence not shown. The network was simulated using NEST [96] and PyNN [97].



SUP. FIG. 6. Dependence of population-averaged input correlations (a), and spike-train correlations (b) on the strength of temporal noise, for the $\text{FB}_{\text{replay}}$ (black circles) and the RAND (light gray circles) case. Symbols and error bars denote mean and standard deviation, respectively, across $M = 30$ network realizations (error bars are partly covered by markers). Gray curve in (a) depicts shared-input correlations in a homogeneous network (Equation 2). Correlations in the FB case correspond to correlations in the $\text{FB}_{\text{replay}}$ case for $\vartheta = 0$ ms. Networks simulated with NEST [96] and PyNN [97].

replaced by a spike time t' randomly drawn from a normal distribution with width ϑ : $t' \sim \mathcal{N}(t, \vartheta)$. Even for small ϑ , correlations increase on the input as well as on the output side, demonstrating the sensitivity of correlations to perturbations in the feedback loop (see Supplements Figure 6 and [5]). This effect is also reflected in an increase of the power of the population activity (see Supplements Figure 7). These results suggest that on hardware temporal noise is responsible for the increase of correlations and population power in the $\text{FB}_{\text{replay}}$ case (compare $\text{FB}_{\text{replay}}$ to FB in Figure 5, 6 and 8).



SUP. FIG. 7. Power spectrum of the population activity of the networks in Supplements Figure 6 for the intact network (solid dark gray), different degrees of dither on input spikes (dashed, dashed dotted and dotted black) and the RAND case (solid light gray), averaged across $M = 30$ network realizations.

Supplements 5: Linear model

We investigate the consistency of our results with a linear rate model that allows us to numerically calculate the average correlations from a given connectivity matrix \mathbf{W} . The model is defined as (according to, e.g., [95])

$$\mathbf{r}(t) = (\mathbf{W}(\mathbf{r} + \mathbf{x}) * h)(t) \quad . \quad (\text{S3})$$

Here, $\mathbf{r}(t)$ denotes the rate of the individual neurons and $\mathbf{x}(t)$ a Gaussian white noise input that is independent for each neuron. The linear filter kernel $h(t)$ depends on the details of the model, is not relevant in our calculation, and hence is not further specified, here. Equation S3 can be transformed to Fourier domain, where the input and output spectral matrices can be expressed by

$$\mathbf{C}_{\text{RR}}(\omega) = T(\omega)T(\omega)^\dagger \quad , \quad (\text{S4})$$

$$\mathbf{C}_{\text{RR}}^{\text{in}}(\omega) = \mathbf{W}\mathbf{C}_{\text{RR}}(\omega)\mathbf{W}^T \quad , \quad (\text{S5})$$

with $T(\omega) = (1 - H(\omega)\mathbf{W})^{-1}$ [95]. In the RAND case, the linear equation for the rate of the (unconnected) neurons reads

$$\mathbf{q}(t) = (\mathbf{W}(\tilde{\mathbf{r}} + \mathbf{x}) * h)(t) \quad , \quad (\text{S6})$$

where $\tilde{\mathbf{r}}(t)$ has the same auto-correlations as $\mathbf{r}(t)$ but zero cross correlations, i.e., $\mathbf{C}_{\tilde{\text{RR}}} = \text{diag}(\mathbf{C}_{\text{RR}})$, since the randomization of spike times removes all spatio-temporal correlations. According to Tetzlaff et al. [5] spectral matrices in the RAND case are given by

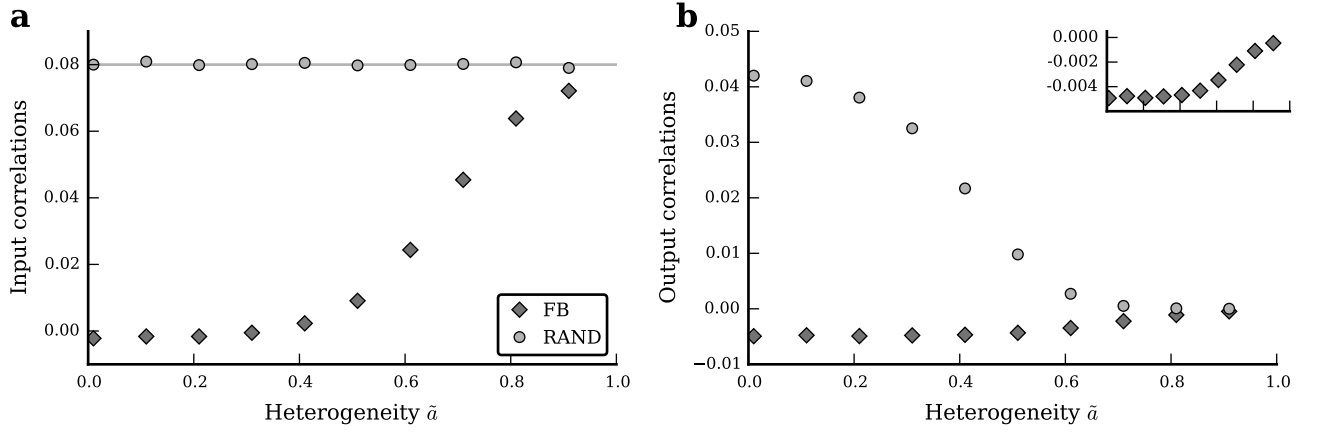
$$\mathbf{C}_{\text{QQ}}^{\text{in}}(\omega) = \mathbf{W}\mathbf{C}_{\tilde{\text{RR}}} \mathbf{W}^T \quad , \quad (\text{S7})$$

$$\mathbf{C}_{\text{QQ}}(\omega) = |H(\omega)|^2 (\mathbf{C}_{\text{QQ}}^{\text{in}} + \rho) \quad . \quad (\text{S8})$$

We calculate the population-averaged power- and cross-spectra from the full matrices:

$$\bar{A}_X(\omega) = \frac{1}{N} \sum_i C_{\text{XX},ii} \quad , \quad (\text{S9})$$

$$\bar{C}_{\text{XX}}(\omega) = \frac{1}{N(N-1)} \sum_{i \neq j} C_{\text{XX},ij} \quad . \quad (\text{S10})$$



SUP. FIG. 8. Dependence of population-averaged input correlations (a), and output correlations (b) on the absolute magnitude of the effective weights (“heterogeneity” \tilde{a} , see text), for the intact network (FB, dark gray diamonds) and the RAND (light gray circles) case. Gray curve in (a) depicts shared-input correlations in a homogeneous network (Equation 2). The inset in (b) shows a magnified view of the spike-train correlations in the FB case (dark gray diamonds). Note that in the linear model the $\text{FB}_{\text{replay}}$ is identical to the FB case, and is hence not shown.

Here, $X \in \{R, Q\}$ denotes the FB and RAND case, respectively. The low frequency coherence is the cross-spectra normalized by the power spectra:

$$\kappa_X(0) = \frac{\bar{C}_{XX}(0)}{\bar{A}_X(0)} . \quad (\text{S11})$$

Note that in the linear model we are actually taking the zero frequency coherence.

As in the spiking model, we consider a sparse network, i.e., we randomly choose for each neuron $i \in [1, N]$ an identical number of presynaptic partners ($K = 15$). In the linear model we do not consider a distribution of non-zero effective weights. Instead, each realized connection is assigned the same weight value $-w$. To mimic the effect of calibration we vary the absolute value of the effective weight by scaling the weights of the non-zero connections with a sigmoidal function of $\tilde{a} \in [0, 1]$:

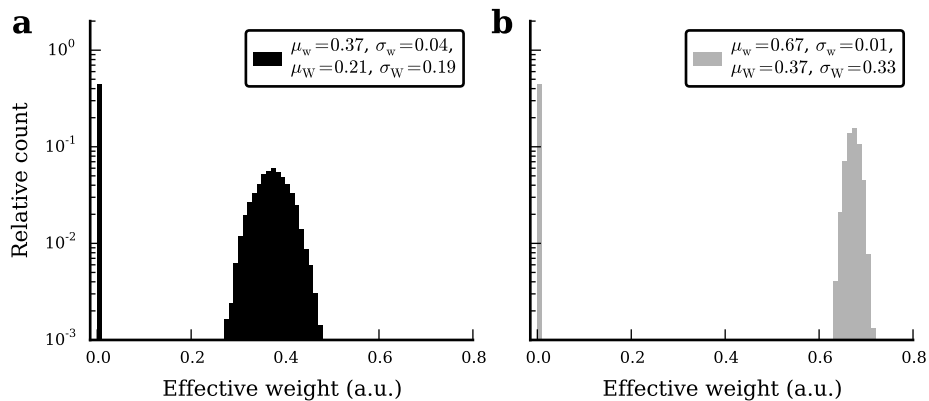
$$\tilde{w} = \frac{1}{1 + e^{10 \times (\tilde{a} - 0.5)}} w . \quad (\text{S12})$$

This procedure changes the variance of the weight matrix [72] and hence \tilde{a} is denoted the *heterogeneity* of the network. More homogeneous (heterogeneous) networks have larger (smaller) effective weights and hence stronger (weaker) feedback. We obtain qualitatively similar results as we observe on the *Spikey* chip (compare Supplements Figure 8 to Figure 8). Correlations in the RAND case decrease, while correlations in the FB case increase with network heterogeneity, i.e., with the variance of the effective weight matrix.

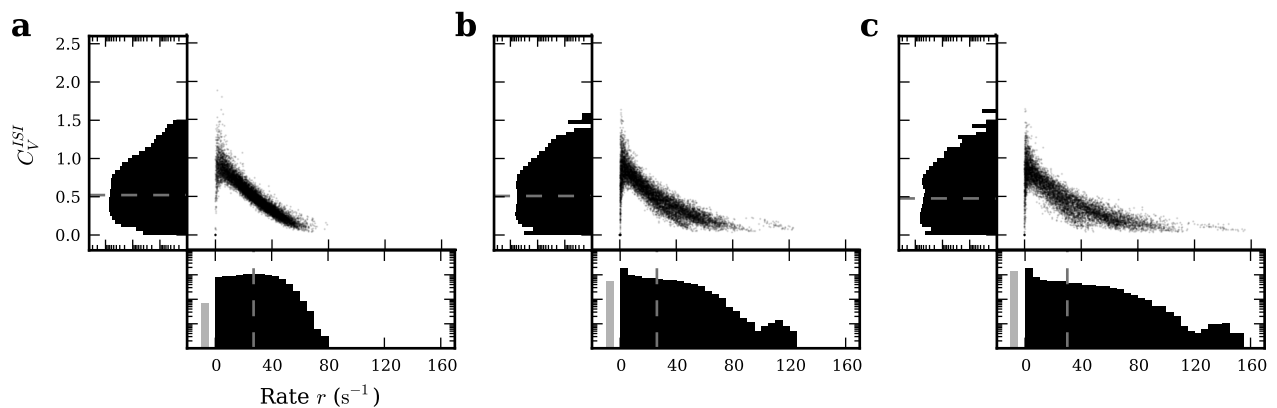
In Supplements Figure 9 we illustrate, how in a sparse network the variance of the weight matrix can increase, although the distribution of non-zero weights becomes narrower. The standard deviation σ_w of a distribution of non-zero weights with mean μ_w is (in this example) smaller than the standard deviation σ_W of the full effective weight matrix, due to the sparseness of the matrix (Supplements Figure 9a; here, we chose $\epsilon = 0.8$). If we, at the same time, increase the mean μ_w and decrease the standard deviation σ_w of non-zero weights, the standard deviation of the weight matrix σ_W can increase significantly (Supplements Figure 9). While the distribution of effective weights is broadened, the mean is decreased, which has a greater impact on the size of correlations. This observation could explain the decrease of correlations with increased calibration of the neuromorphic chip.

Supplements 6: Firing statistics in the RAND case

The firing rate distributions in the RAND case are similar to those in the FB scenario (compare Supplements Figure 10 to Figure 7d–f). They are narrower for more homogeneous networks. Nevertheless, former inactive neurons in the FB case have a higher probability to fire in the RAND case, because the temporal fluctuations of their membrane potentials increase due to higher correlations in their input (Figure 5e). Neurons with firing rates above average are



SUP. FIG. 9. Distributions of effective weights in a sparse network. Different values for the mean μ_w and standard deviation σ_w of the distribution of non-zero weights in (a) and (b), respectively, result in different values for the mean μ_W and standard deviation σ_W of the effective weight matrix.



SUP. FIG. 10. (a–c) Like Figure 7d–f, but for the RAND case. The mean of firing rate (27.2 s^{-1} , 26.0 s^{-1} , 30.0 s^{-1}) and C_V^{ISI} distributions (0.52 , 0.51 , 0.48) are marked with dashed lines.

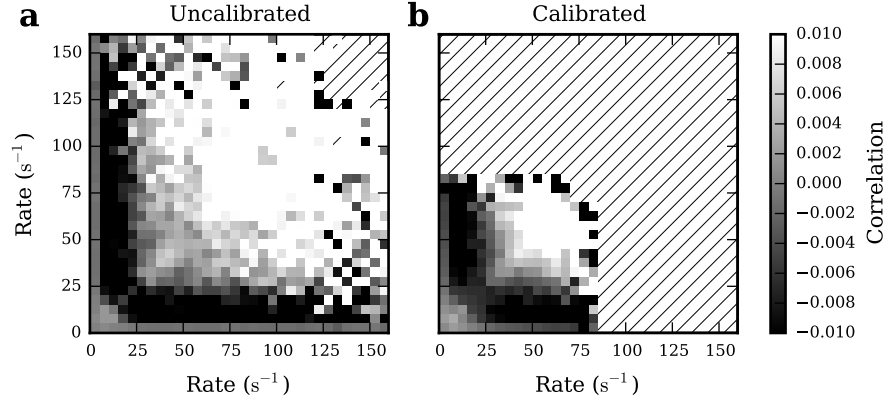
barely affected by this effect, because they are strongly driven by the constant current influx, and hence show similar firing rates than in the FB scenario. The regularity of firing increases for the RAND compared to the FB case, also due to stronger fluctuations of the input.

Supplements 7: Correlation matrix

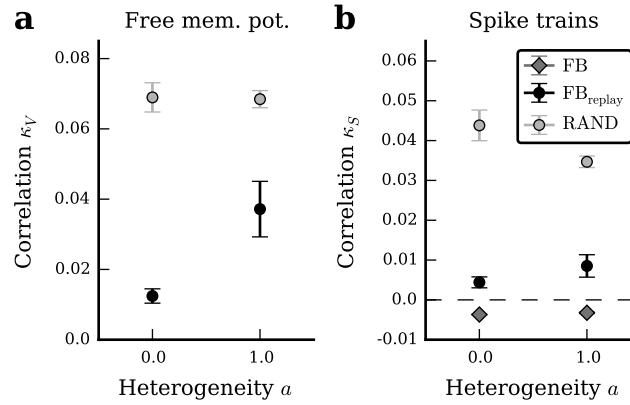
In addition to the population-averaged measures from the main manuscript, we also calculated the pairwise correlations for each pair i, j of neurons with $i \neq j \in [1, N]$, and ordered these by the time-averaged rate of the corresponding neurons (Supplements Figure 11). This reveals a dependence of the pairwise correlation on the rate of the respective neurons. If both neurons fire at low rate (here $< 5\text{ s}^{-1}$), correlations will be close to zero similar to the results in [89]. For high rates (here $> 25\text{ s}^{-1}$) we find mostly positive correlations. However, if both neurons fire at intermediate rates, the activity of neurons will be anti-correlated, and hence suppresses positive shared-input correlations. After calibration, the amount of neurons firing at intermediate rates increases, and hence shared-input correlations are suppressed by more neurons (Supplements Figure 11, Figure 7).

Supplements 8: Results for different *Spikey* chips

The experimental protocol presented in the main text was used for two additional *Spikey* chips. Different chips show different realizations of fixed-pattern noise, and hence calibration was repeated for each chip separately (Figure 4,



SUP. FIG. 11. Pairwise spike-train correlations for all pairs of neurons in $M = 100$ network realizations, sorted by the rate of the respective neurons. Diagonal stripes indicate that no data was available. Here $C_{ij}(f) \in \mathbb{C}$, but we only consider the real part of the cross spectrum to calculate the low-frequency coherence: $\kappa_{ij} = \frac{1}{f_{\max} - f_{\min}} \int_{f_{\min}}^{f_{\max}} \frac{\Re(C(f))}{A(f)}$.

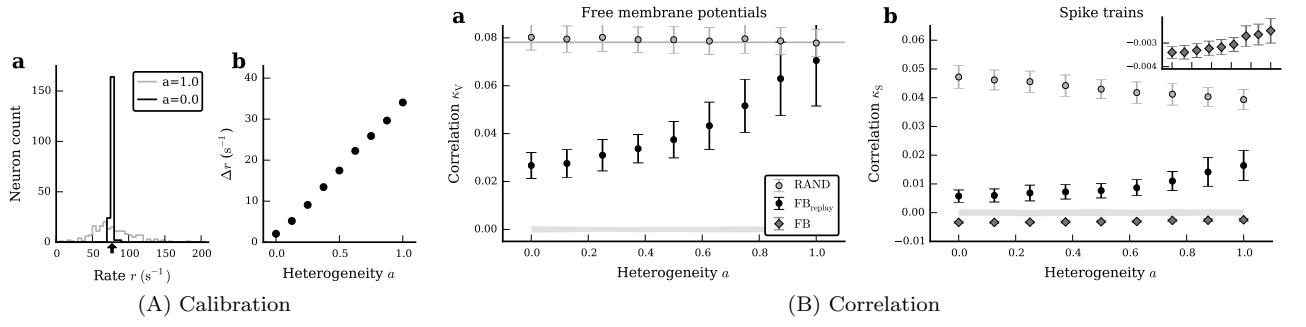


SUP. FIG. 12. Like Figure 8 and for the same chip, but for $L = 20$ trials of one network realization.

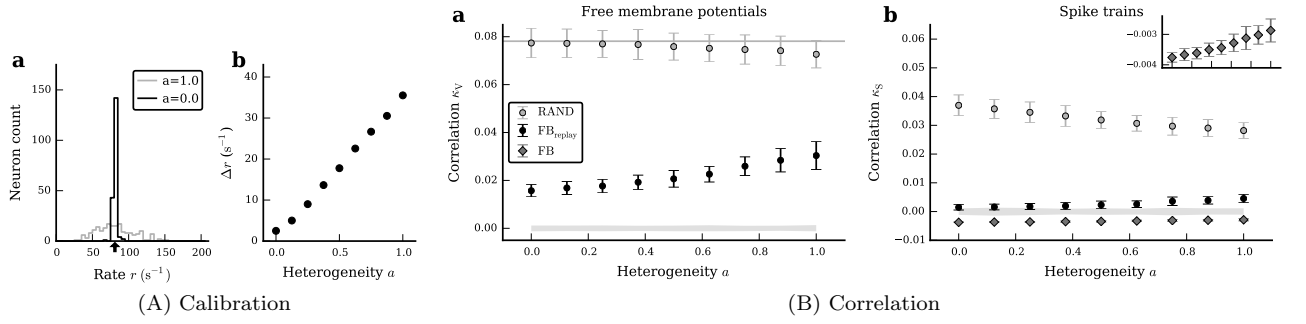
Supplements Figure 13A and 14A). In the calibrated state free membrane potentials (and spike trains) are decorrelated by inhibitory feedback for all chips (Figure 8, Supplements Figure 13B and 14B). However, the more uncalibrated the system is, the more the results differ between chips, which is likely to be caused by different extents of intrinsic fixed-pattern noise. This is most pronounced for chip 2 (Supplements Figure 13B), where the input correlations of the $\text{FB}_{\text{replay}}$ case reaches that of the RAND scenario for the uncalibrated state, which means that the input of neurons is not decorrelated by the inhibitory feedback anymore.

Supplements 9: Reproducibility of networks with intact feedback

We measured the variance of free membrane potential and spike train correlations over several trials for a single network realization (Supplements Figure 12). This variance is smaller than the variance we observe over different network realizations (compare to Figure 8 and Supplements Figure 13 and 14), which indicates that the latter is mostly caused by the different connectivity, not by trial-to-trial variability. Note that the variability between trials of networks with intact feedback is likely to be larger than between replays of network activity as shown in Figure 3, because network dynamics may be chaotic. The data shown in Supplements Figure 12 has to be interpreted with care, because the reproducibility of only a single network realization is considered. For different realizations, the variance may change.



SUP. FIG. 13. Like chip 1 in Figure 4 and 8, but for chip 2 in (A) and (B), respectively.



SUP. FIG. 14. Like Supplements Figure 13, but for chip 3.

A			Model summary								
Populations			One (inhibitory)								
Topology			-								
Connectivity			Random convergent connections (fixed in-degree)								
Neuron model			Leaky integrate-and-fire (LIF), fixed firing threshold, fixed absolute refractory time								
Channel models			-								
Synapse model			Exponentially decaying currents, fixed delays								
Plasticity			-								
External input			Resting potential higher than threshold (= constant current) ($E_l > \Theta$)								
Measurements			Spikes and membrane potentials								
Other			No autapses, no multapses								
B						Populations					
Name			Elements			Size					
I			LIF neuron			N					
C						Connectivity					
Source		Target		Pattern							
I		I		Random convergent connect, in-degree K							
D						Neuron and synapse model					
Type			Leaky integrate-and-fire, exponential currents								
Subthreshold dynamics			Subthreshold dynamics ($t \notin (t^*, t^* + \tau_{\text{ref}})$): $C_m \frac{d}{dt} v(t) = -g_l(v(t) - E_l) + I_{\text{syn}}(t)$ Reset and refractoriness ($t \in (t^*, t^* + \tau_{\text{ref}})$): $v(t) = v_{\text{reset}}$								
Current dynamics			$\tau_{\text{syn}} \frac{d}{dt} I_{\text{syn}}(t) = -I_{\text{syn}}(t) + \sum_{i,k} J \delta(t - t_i^k)$ Here the sum over i runs over all presynaptic neurons and the sum over k over all spike times of the respective neuron i								
Spiking			If $v(t^*-) < \Theta \wedge v(t^*+) \geq \Theta$: emit spike with time stamp t^*								
E						Measurements					
Spike trains			recorded from P_s neurons								
Membrane potentials			recorded from P_v neurons								

SUP. TABLE I. Description of the network model (according to [104]).

B Populations		
Name	Values	Description
N	$\{192, 4800\}$	network size
C Connectivity		
Name	Values	Description
K	$\{15, 375\}$	number of presynaptic partners
D Neuron		
Name	Values	Description
C_m	0.2 nF	membrane capacitance
τ_{ref}	0.1 ms	refractory period
v_{reset}	-80 mV	reset potential
E_l	-52 mV	resting potential
Θ	$\sim \mathcal{N}(-62, [0, 8.8])$ mV	firing threshold
g_l	10 nS	leak conductance
D Synapse		
Name	Values	Description
τ_{syn}	5 ms	synaptic time constant
J	-0.254 nA	synaptic weight
d	1.0 ms	synaptic delay
E Measurements		
Name	Values	Description
P_s	$\{192, 4800\}$	number of neurons spike trains are recorded from
P_v	$\{150, 150\}$	number of neurons membrane potentials are recorded from

SUP. TABLE II. Parameter values for the network model described in Supplements Table I with distributed thresholds.

B Populations		
Name	Values	Description
N	192	network size
C Connectivity		
Name	Values	Description
K	15	number of presynaptic partners
D Neuron		
Name	Values	Description
C_m	0.2 nF	membrane capacitance
τ_{ref}	0.1 ms	refractory period
v_{reset}	−80 mV	reset potential
E_l	−52 mV	resting potential
Θ	−62 mV	firing threshold
g_l	10 nS	leak conductance
D Synapse		
Name	Values	Description
τ_{syn}	5 ms	synaptic time constant
J	$\sim [\mathcal{N}(0.254, [0.00254, 0.75])]_+$	synaptic weight, clipped to positive values
d	1.0 ms	synaptic delay
E Measurements		
Name	Values	Description
P_s	192	number of neurons spike trains are recorded from
P_v	150	number of neurons membrane potentials are recorded from

SUP. TABLE III. Parameter values for the network model described in Supplements Table I with distributed weights.

A			Model summary								
Populations			Two (excitatory, inhibitory)								
Topology			-								
Connectivity			Random convergent connections (fixed in-degree)								
Neuron model			Leaky integrate-and-fire (LIF), fixed firing threshold, fixed absolute refractory time								
Channel models			-								
Synapse model			Exponentially decaying currents, fixed delays								
Plasticity			-								
External input			Resting potential higher than threshold (= constant current) ($E_{\text{l}} > \Theta$)								
Measurements			Spikes and membrane potentials								
Other			No autapses, no multapses								
B						Populations					
Name			Elements			Size					
E			LIF neuron			N_{E}					
I			LIF neuron			N_{I}					
C						Connectivity					
Source		Target		Pattern							
E		E		Random convergent connect, in-degree K_{E} , weight J_{E}							
E		I		Random convergent connect, in-degree K_{E} , weight J_{E}							
I		E		Random convergent connect, in-degree K_{I} , weight J_{I}							
I		I		Random convergent connect, in-degree K_{I} , weight J_{I}							
D						Neuron and synapse model					
Type			Leaky integrate-and-fire, exponential currents								
Subthreshold dynamics			Subthreshold dynamics ($t \notin (t^*, t^* + \tau_{\text{ref}})$): $C_{\text{m}} \frac{\text{d}}{\text{d}t} v(t) = -g_{\text{l}}(v(t) - E_{\text{l}}) + I_{\text{syn}}(t)$ Reset and refractoriness ($t \in (t^*, t^* + \tau_{\text{ref}})$): $v(t) = v_{\text{reset}}$								
Current dynamics			$\tau_{\text{syn}} \frac{\text{d}}{\text{d}t} I_{\text{syn}}(t) = -I_{\text{syn}}(t) + \sum_{i,k} J \delta(t - t_i^k)$ Here the sum over i runs over all presynaptic neurons and the sum over k over all spike times of the respective neuron i								
Spiking			If $v(t^* -) < \Theta \wedge v(t^* +) \geq \Theta$: emit spike with time stamp t^*								
E						Measurements					
Spike trains			recorded from P_{s}^{E} excitatory and P_{s}^{I} inhibitory neurons								
Membrane potentials			recorded from P_{v}^{E} excitatory and P_{v}^{I} inhibitory neurons								

SUP. TABLE IV. Description of the network model consisting of an excitatory and an inhibitory population (according to [104]).

B Populations		
Name	Values	Description
N_E	96	size of the excitatory population
N_I	96	size of the inhibitory population
C Connectivity		
Name	Values	Description
K_E	7	number of excitatory presynaptic partners
K_I	8	number of inhibitory presynaptic partners
D Neuron		
Name	Values	Description
C_m	0.2 nF	membrane capacitance
τ_{ref}	0.1 ms	refractory period
v_{reset}	−80 mV	reset potential
E_l	−52 mV	resting potential
Θ	−62 mV	firing threshold
g_l	10 nS	leak conductance
D Synapse		
Name	Values	Description
τ_{syn}	5 ms	synaptic time constant
J_E	0.0635 nA	excitatory synaptic weight
J_I	−0.254 nA	inhibitory synaptic weight
d	1.0 ms	synaptic delay
E Measurements		
Name	Values	Description
P_s^E	96	number of excitatory neurons spike trains are recorded from
P_s^I	96	number of inhibitory neurons spike trains are recorded from
P_v^E	150	number of excitatory neurons membrane potentials are recorded from
P_v^I	150	number of inhibitory neurons membrane potentials are recorded from

SUP. TABLE V. Parameter values for the network model described in Supplements Table IV.